Volume 26 Number 1 March 1975

BJPIA5 26 (No. 1) 1-90 (1975) 0007-0882

The British Journal for the Philosophy of Science

17

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editors: JOHN WATKINS and JOHN WORRALL

Assistant Editors: COLIN HOWSON and ELIE ZAHAR

Subscriptions
The *Journal* is published quarterly. The subscription price is £7.00 net (U.S.A. and Canada U.S. $22.00) per volume post free; single issues £2.50 net (U.S.A. and Canada U.S. $8.00) plus postage. The subscription is £4.50 (U.S.A. and Canada U.S. $14.00) for members of the British Society for the Philosophy of Science, who should order direct from the publisher. (Membership of the Society is international and is by no means confined to people resident in the United Kingdom.) Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

Notes to Contributors
Articles, correspondence and books for review should be sent to John Worrall at the Department of Philosophy, London School of Economics, Houghton Street, Aldwych, London WC2A 2AE. Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Footnotes, which should be as few as possible, should be typed separately, double-spaced at the end of the text. They should be referred to by consecutive numbering throughout the text, in which they will be inserted in page proof. Bibliographical references should be listed alphabetically at the end. The style of references should follow current *Journal* practice.

Long quotations are printed in small type without quotes, and should be so marked in the typescript. Single quotes should be used to form names of words and for shorter quotations; double quotes for other purposes. Diagrams, tables, and illustrations should be on separate sheets, with their desired position in the text clearly indicated. Symbolic formulae appearing as separate lines should be clearly distinguished in the text, and foreign letters (Greek, German, etc.) indicated in the margin. Material to be printed in italics, especially English letters used as symbols ('*A*', '*x*', etc.), should be underlined. On matters of style and layout, authors should consult CHAUNDY, T. W. *et al.* [1964]: *The Printing of Mathematics*, London: Oxford University Press.

Offprints of articles may be ordered when corrected galley proofs are returned. Fifty are supplied free ; the price of extra copies is given on the order form.

© British Society for the Philosophy of Science 1975

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editors.

Back Volumes
Volumes 1-16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17-25 may be obtained from Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Prices on application.

CUK-H00666-17-P8869

Brit. J. Phil. Sci. 26 (1975), 1–18 *Printed in Great Britain*

501.05
B774

Imre Lakatos*

by PAUL FEYERABEND

1 Imre Lakatos, who died suddenly and quite unexpectedly in February 1974, was a fascinating person, an outstanding thinker and the best philosopher of science of our strange and uncomfortable century. He was a rationalist, for he thought that man had the duty of using reason in his private affairs as well as in any inquiry concerning the relation between himself, nature, and his fellow man. He was an optimist for he thought that reason was capable of solving most of the problems arising in the course of such an inquiry. He had realistic ideas about this capability of reason for he emphasised that it cannot always be expressed in the form of rules, and that a mere abstract comparison of rules is quite insufficient for improving it even if the comparison should be guided by general principles of the most profound nature: if reason is to have a point of attack in this world with its complicated episodes and its hair-raising ideas and institutions, then it must be both sly and sophisticated. In the sciences, for example, reason must not be more primitive than the theories which it is supposed to evaluate; but reason must not be too strict either, otherwise it becomes impossible to build up science or else to improve ideas, experiments, institutions which are already in existence. We need a type of reason that keeps a sane balance between the inactive (though by no means silent) *contemplation* of science, and reformatory zeal. The demands of reason have to be adapted to the (historical, psychological, physical, financial) properties of the material to be changed.

This side of methodology has hardly ever been examined by philosophers of science. The standard procedure here is first to invent a very simple-minded model of knowledge and then to evaluate rules by comparing them with the model. Popper's theory of falsification, for example, arose from the trivial observation that while a singular statement may entail the negation of a universal statement it never entails any universal statement. One soon realised that the suggested rules were either useless or

---

[*Editors' Note*]: Imre Lakatos, then Editor of this *Journal*, died suddenly on 2 February 1974. Shortly afterwards Professor Feyerabend was invited by the then Acting Editor to write an appreciation of Professor Lakatos.

[*] An earlier version of this article was extensively criticised by Professor J. W. N. Watkins and Mr John Worrall. I have accepted some of their suggestions.

devastating. The material of the scientist is much too incoherent and unwieldly to survive an out and out Popperian attack.

On the other hand, this material is not beyond the reach of every epistemological criticism. Analysing methods of mathematical proof,[1] Imre Lakatos found that complicated theories are often attempts to cover up difficulties that arise in the course of a discussion of simple and even quasi-empirical problems. Some problems are solved, and additional problems that arise in the course of this solution are covered up by complicated machinery. On the other hand it is possible to suggest solutions which lead to new and unexpected problems and which make us discover new and unexpected (mathematical) facts. If theories of the latter kind are preferred despite all the difficulties they create, if they are given time to develop then one has a criterion that is strict without being fatal, that has a point of attack in scientific practice without restricting the freedom of the scientist, the freedom of daring intellectual enterprises included. Lakatos has extended the criterion to the empirical sciences and has explained it in detail in his epoch making essay 'Falsification and the Methodology of Scientific Research Programmes'.[2] The theory of science that underlies the criterion combines strict criticism with free decision, historical accident with rules of reason. It is one of the most important achievements of twentieth-century philosophy. It is important because it solves almost all the problems created by the first rise of Western Rationalism (the Presocratics) and by the repeat performance in the sixteenth and seventeenth centuries.

But Imre Lakatos was not merely a theoretician. He was also a crusader, journalist, politician. He combined arguments with brilliant performances during his lectures, biting humour in discussions, charming persuasiveness in private conversation. Very soon there arose a circle of people of similar intellectual tastes—I advisedly do not use the word 'thinkers', for mere thinking-machines were a horror to Lakatos—who were interested in removing that endless and conceited boredom which today seems to be the prime characteristic of scholarship. The circle was not restricted to rationalists, for Imre Lakatos was anything but narrowminded, and even a Dadaist like myself could work with him and profit from his advice. Then students and older colleagues joined this new intellectual community, founded on friendship and sceptical interest rather than on slavish subservience. This was possible, for Imre Lakatos was above all a kind and warmhearted human being, deeply concerned about the increasing irrationality and injustice in this world, about the almost insurmountable power which mediocrity possesses in all fields and among all people,

---

[1] Lakatos [1963–4].

[2] Lakatos [1970].

P869

including the so-called 'youth'; prepared to defend his ideas and his attitude even under uncomfortable and dangerous circumstances; not to be silenced and yet not without being aware of the ultimate absurdity of all human enterprise. A man like him can never be completely understood on the basis of his written work alone. And this work will survive only in the hands of those who possess similar freedom, similar joy of life, and a similar abundance of ideas.

My own evaluation of Imre Lakatos's work will be decidedly subjective. I shall look at some aspects of his theory in a rather personal way, as one looks at the work of a friend and *not* as one looks at the work of a faceless scholar of an age long gone by. I shall emphasise elements which are not apparent from the published accounts and whose existence is denied by his more rationalist colleagues. I shall also give a brief sketch of some historical developments which at first sight don't seem to have anything to do with the methodology of research programmes. I do this because I believe that these developments have created problems which were not solved before but which Imre Lakatos has brought much closer to a solution. I start my historical sketch with an account of the Presocratics.

2 Science started in the West when some extraordinary people preferred the results of certain intellectual games to what they could learn from their senses and from tradition. (When I speak of 'games' rather than of 'rules of thinking' or 'intellectual discoveries' I do this to emphasise the often quite *arbitrary* character of the principles then introduced. The Presocratics only rarely argued with their predecessors, they started a new form of life and measure of excellence.[1]) We do not know how the games that constituted this measure came into being and we can only be surprised at the way in which they soon affected perception, thought, social relations. Epistemology and the philosophy of science arose from the many difficulties created by this development. *Xenophanes*, for example, had pointed out that God cannot be many and that he cannot be created. If he is many, then the many are either equal or unequal; if equal, then they are one, if unequal, then those different from the one that *is* cannot *be*, so there will again be only one. On the other hand, if God is created, then he comes either from the same, or from what is different. If he comes from the same, then he is not created; but the different is what is not, and nothing can come from what is not. These arguments contradicted both *tradition* (Homer, Hesiod) and *direct experience* (which at the time

---

[1] For example, the older thinkers are criticised for their *πολυμαθίη* but nobody explains why a unified account is preferable to a list.

in question included the experience of divine intervention brought about by a great variety of partly helpful, partly mischievous gods),[1] they introduced entirely new tales claiming to contain superior truth, and they called for a new kind of investigation dealing with the relation between the tales and the conventional accounts.

Starting the new kind of investigation in a straightforward and rather dogmatic manner Xenophanes declares that his 'arguments' overrule whatever is in conflict with them. The feeling, so characteristic for the early Greeks, that nature is the playground of the gods and that it is full of intelligent life *is mistaken*, even if it should find support in experience and language.[2] On the other hand, it is possible to *reform experience* by eliminating its theological ingredients. Such a reformed experience no longer tells us 'that the sun rises' (speaking in this way one assumes a permanence for which no reason can be found, which is suspiciously close to the older theology and which, moreover, is contradicted by phenomena such as the increasing shapelessness of 'the sun' as it approaches haze and clouds on the horizon), or that there is a sky (the idea of a sky is a remnant of older views[3] found unacceptable by the argument). Nor is this reformed experience a natural process. It is an *artifact*, both phenomenologically (the observer has to be *trained* not to project too much life into the external world; he must *learn* to see the sun as a stone and the sky as a complicated mechanical system) and theoretically (he must not use the wrong concepts either; for example, he must no longer use the customary concepts of hot and cold). Introducing this artifact into cosmological speculation Xenophanes creates empiricism with its dichotomy between *evidence* (parts of the new type of experience just invented) and views that go beyond it but are in need of *evidential support*, and the quite different dichtomy between our natural experience and the manufactured processes that alone can count as evidence. And we have now the problem how the right kind of support can be found, *i.e.* we have now a *problem of evidence* and a *problem of induction*.[4]

Xenophanes's 'proof' also shows that cosmologies that explain how the world arose from simpler forms are basically incorrect. We can say *what the world is*, we cannot say *how it has become what it is*. *Historical accounts* such as those of Hesiod, and of Anaximander must therefore be replaced by recitation of basic principles. Professor Szabó has shown[5] how this leads directly to the *axiomatic approach* which was to dominate western thought down to our own times.

---

[1] Nietzsche [1873]; Wilamowitz-Moellendorf [1955]; Dodds [1957]; Nilsson [1949] *etc.*

[2] Terms such as 'hot' and 'cold' included the heat of love and the coldness of anger and hatred.

[3] Such as the old idea of a separation of heaven and earth.

[4] This is a first approximation only.

[5] Szabó [1969].

Aristotle proceeds on the path that had been opened by Xenophanes and Parmenides, but he modifies it in a most interesting way. He retains the idea that knowledge proceeds from principles. He also retains the first of the two dichotomies mentioned above but he does not retain the second one.[1] He returns to the earlier view that man's natural experience and his natural ways of talking are basically correct and he tries to harmonise them with the newly found machinery of abstract argument. He knows quite well that our senses are not perfect and he discusses numerous instances where error is likely to occur. But there are conditions and contents of perception that can be trusted. It is these conditions and these contents that lead to the principles of our sciences as well as to a general theory of motion and causation. They lead to them in two ways, logically, by excluding extravagant views (Aristotle has a principle of falsification) and causally, by giving rise to the ideas that constitute the principles (evidence and principles are manufactured by the organism itself who *receives* the forms inherent in nature, *retains* them through a process of gradual 'imprinting' and finally *imposes* them on material more chaotic than what happens under normal conditions, leading to clear perceptions even in this case: the problem of evidence and the problem of induction disappear from philosophy and are 'solved' in physiology). Perceptions are not changed and castrated in the way suggested by Xenophanes. But they are not regarded as being given either. Each single perceptual occurrence can be explained on the basis of a general theory of motion that comprises locomotion, sensation, biological growth *etc.* and that indicates when and under what physical circumstances error is going to arise (perceptual error is not a result of wilfulness, it is simply a case of physical malfunctioning).

This fantastic and unique cosmology which gives abstract speculation its due without permitting it to throw doubt on the basic abilities of man and which makes knowledge part of the physical universe was overwhelmed and pushed aside by the *rise of modern science.* In the past few years it has been made clear (by Burtt, Koyré and others) that modern science did not arise because of empirical problems created by the inadequacy of earlier forms of thought; it arose because *a new ideology was imposed from the outside.*[2] In a way the situation was identical with what happened at the time of the Presocratics but the lines were not drawn as clearly as in the earlier instance. The proponents of the new views were not just content to present their ideas and to let history take its course, they tried to get support from the opposition as well, and so they embarked on one of the most successful propaganda campaigns in the history of thought. In consequence the new cosmology appeared less radical and commonsense more sophisticated than they actually were. As modern science developed one realised that the following conflicts with the Aristotelian point of view had arisen.

---

[1] With Parmenides, Zeno and Plato other problems are added to those two: Western rationalism starts with an attempt to solve the problems it has created by its appearance (and which were absent from the earlier world views).

[2] *Cf.* for example Gingerich [1974]. For a fascinating account of how the new views gradually crept in, insinuated themselves and so took over without argument *cf.* Kocher [1953].

(1) Modern science contradicts Aristotle's claim that natural man is in harmony with the world in the sense that his natural powers (sense perception; language *etc.*) do not deceive him about its basic structure; (2) it contradicts Aristotle's claim that the basic principles are, and must be, necessary; (3) it uses a new theory of motion that is restricted to locomotion promising that other forms of motion will eventually be reduced to it, but it never keeps the promise (this procedure and this development are partly responsible for the renewed appearance of a gulf between 'the evidence of our senses' and the creatures of pure thinking); (4) modern science also brings with it a new *aim* for scientific research. *Before* the aim had been to develop a cosmology that could in principle accommodate any fact. Failure to accommodate a fact indicated either that the correct cosmology had not yet been found, or threw doubts on the methods used in the discovery of the fact. *Now* the constant expansion of the horizon of facts becomes a necessary condition of acceptable science. The new aim is realised only very gradually and we are not too far from the truth when saying that the first *explicit* formulation of it is found in Lakatos (or, if one wants to quibble, in Popper, as reformulated by Lakatos).

3 It is clear that radical changes such as these could not be explained by the traditional epistemologies that often excluded them, and new views were proposed to cope with the situation. The development that had started with the Presocratics and that had given rise to so many problems now continued, after a temporary halt. In a way we are still in the middle of this development, we are still trying to understand science and the kind of rationalism that was invented by the Greeks. This is the reason why I have started my sketch with Xenophanes. The new views were either comprehensive, or more restricted. The more restricted views describe an ideal of knowledge and indicate how science fits the idea. The comprehensive views give an account of the changes that are needed to transform Aristotelianism, or any other non-scientific ideology into (modern) science. Most of the time the ideal is taken for granted, it is

not discussed (this applies even to a sophisticated thinker such as Imre Lakatos).

Among the comprehensive views the most interesting and most influential one is that of *Bacon*. Bacon realises that a new science must be based on a new experience and on new ways of thinking. And as a new experience and new ways of thinking can only arise in a new man, he wants to rebuild man so that he may become a plain and undistorting mirror of the world.[1] The activity that brings about this great reconstruction cannot be based on experience: the old experience is worthless, the new experience has not yet arrived. The activity is therefore without a foundation: Bacon is the first modern thinker to conceive the idea of *a science without foundations* (but he ascribes to it a transitory function only: it leads up to a future science that again rests upon a foundation, but this time the foundation is trustworthy, and agrees with the world).

For a long time *Newton's science* was thought to be the new and well founded science that Bacon had promised. The 'phenomena' on which it rests are neither sensations, nor normal experiences. They are simple laws of nature, somewhat idealised, expressed in mathematical terms, and occasionally capable of direct experimental demonstration.[2] Newton also developed a methodology and philosophy of science. It belongs to the restricted type and has been the model for almost all later thinking on the matter. According to Newton, science is built up by first finding phenomena, then deriving laws, then generalising these laws, always testing the generalisations with the help of new experiments, and by finally framing hypotheses for their explanation. Phenomena, laws, hypotheses are corrigible by further research. For example, a law that has been obtained by generalising a statement derived from phenomena may be restricted as a result of further observations. But such corrections can only come from the same level, or from a lower level; they cannot come 'from above'. For example, hypotheses may not correct phenomena.

Newton's faith in having found the correct method for discovering and justifying general statements was supported by his apparent success in deriving his law of gravitation 'from phenomena'. Newtonianism dominated science and the philosophy of science down to the twentieth century.[3] Of course, *Hume* showed that justification can be neither by derivation from experience nor by probabilification—but his objections had only little immediate influence. They were not formulated in mathematical terms and, besides, they seemed to deny Newton's achievements, or turn them into a series of accidents. Nor was *Hegel* more successful when he compared Newton's theory of perturbations with his 'derivation' pointing out that the latter was a mere transcription of Kepler's formulae.[1] And even in the twentieth century, and after *Duhem*, the effort was continued to find ways of using evidence in a 'positive' way, *i.e.* to define methods that would select theories, without any interference 'from above' by showing that they were 'well supported' by the evidence, or at least better supported than their rivals.[2] There were only two philosophies that tried to give a different account of the relation between evidence and theories, Kant's transcendental idealism, and the philosophy of conventionalism.

---

[1] Bacon [1608], aphorisms 41, 50, 95, 115. When Bacon speaks of 'unprejudiced senses' (97) he does not mean sense data, or 'raw feels'; he means the perceptions of the man of the future who is attuned to reality.

[2] *Cf.* my [1970].

[3] *Cf.* Born [1949].

Conventionalism agrees that theories cannot be 'justified' in the way imagined by simple forms of inductivism. They are justified because of the order they bring into the known facts and because they provide concepts and ordering principles for things as yet to be discovered. The order is never complete, for there are always recalcitrant phenomena. This does not invalidate the chosen scheme, but challenges its defenders to rebuild the phenomena until they fit into it. The scheme is chosen either because it easily accounts for some empirical regularities (empirically motivated conventionalism), or because it follows from certain theoretical postulates (Dinglerian version). Once the scheme is accepted and incorporated into scientific practice, it quite automatically transforms the facts and thus removes any criticism 'on the basis of experience'. *Transcendental idealism* assumes that man classifies data automatically, without having made a conscious decision to adopt an ordering scheme. The laws which underlie his classifications are found and proved, by transcendental deduction.

Neither transcendental idealism nor conventionalism (of the empirically motivated kind) are universal theories. They offer means of justification for some very general views, but they become inductivist when dealing with simple regularities. And even in the case of the preferred views, the solution depends on empirical laws: Kant assumes that the human mind does not change, the conventionalists (and this now includes the Dinglerians) assume that the decision of the scientist to stick to certain forms of thought will keep these forms of thought imprinted on his material forever. The problem of justification moves from one place to another, it is not solved.

---

[1] Hegel [1817], section 270.

[2] For the strange story of the theories that arose in this way and their progressive degeneration *cf.* Lakatos [1968]. Duhem's result, incidentally, when translated into the lower predicate calculus reveals the need for a logic of confirmation where 'all ravens are black' can be highly confirmed by grey ravens.

4 It remains unsolved in yet another theory that was formulated twice in the nineteenth century, first by Mill, in his immortal essay *On Liberty*, and then by thinkers such as Boltzmann who extended Darwinism to the running of science itself. This theory takes the bull by the horns: theories *cannot* be justified, and they *need not* be justified. *There is no positive explanation of the success of scientific hypotheses, theories, laws, except by means of other hypotheses, theories, laws. And there are no positive reasons for accepting hypotheses, theories, laws.* We cannot derive them from facts, we cannot guarantee that they will remain successful after their first successes, we cannot guarantee that the facts are as described, we cannot even guarantee that the theories we have chosen are in agreement with the demands of formal logic. But all that is not necessary. The procedure of science is not to praise the good, but to eliminate what is bad, or comparatively inferior. The 'fundamental difference between [this] approach and [inductivism, or other types of justificationism] is that [it] lays stress on *negative arguments*, such as negative instances, or counter-examples, refutations and attempted refutations—in short criticism' while empiricists of the Newtonian tradition usually emphasise positive arguments, positive instances and the like.[1] 'So essential is this [critical] discussion to a real understanding of moral and human subjects that, if opponents of all important truth do not exist, it is indispensable to imagine them and to supply them with the strongest arguments which the most skillful devil's advocate can conjure up'.[2] The theory that survives the debate is of course not free from fault. It dominates the field not because it has been proved (or because it is highly probable), or because it possesses some other desirable feature such as consistency. Even the best scientific theory is usually in a pretty sorry state. What *is* decisive is that its faults are smaller than the faults of the rivals—for the time being. Thus 'the beliefs which we have most warrant for have no safeguard to rest on but a standing invitation to the whole world to prove them unfounded' that is, to find reasons for rejecting them. 'If we use every opportunity for contesting it (a certain hypothesis) has not been refuted' then we shall regard it as better than an alternative which, while supported by all the available evidence 'has not gone through a similar process'. 'If even the Newtonian philosophy were not permitted to be questioned, mankind could not feel as complete assurance of its truth as they do now'.[3] Thus 'the edifice of our knowledge does not consist of . . . well established truths . . . It consists of largely arbitrary elements . . . so-called hypotheses'.[4]

[1] Popper [1972], p. 20. Popper, of course, says 'my' instead of 'this'.

[2] Mill [1859], quoted from Cohen [1961], p. 228. Mill does not restrict his method to 'moral and human subjects', he applies it to scientific theories as well. See *below*.

[3] Mill, *op. cit.* pp. 209, 208, again 209.

[4] Boltzmann [1905], chapter 8.

Or, in Popper's repetition: '. . . all that can possibly be '*positive*' in our scientific knowledge is positive *only* insofar as certain theories are, at a certain moment of time, preferred to others in the light of our *critical* discussion. . . . Thus even what may be called 'positive' is so *only* with respect to *negative methods*'.[1]

The analogy with *Darwin* is obvious. Before Darwin it was customary to view organisms as divinely created and *therefore perfect* solutions to the problems of life on this planet. During the nineteenth century one became aware of the numerous failures of the life process. This process is not a carefully planned and meticulously performed attempt to realise an aim that has been thought out in advance. It is unreasonable, wasteful, it produces an immense variety of forms *and leaves it to nature to eliminate the duds*. The remaining forms are surprisingly efficient, *as if* they had been planned with a definite aim in mind. However, there was no aim, there was no 'method of justification', and it is doubtful whether conscious planning could ever have produced a comparable result.

The analogy does not *solve* methodological problems, it only shows what kinds of solution *are possible*. The world of the organism is a natural world, it acts blindly. The world of theory is a social world, it is built up by scientists *who have to decide* what to keep and what to eliminate. Is the decision to be completely arbitrary, is it supposed to proceed according to explicit rules and, if the latter, which rules shall we choose? These are the questions which arise once we start relying on methods of elimination instead of looking for methods of justification.

Mill dispenses with general rules and judges every case on its own merit, and in the concrete historical situation in which it occurs.[2] He seems to think that the shape of the rival theories and the historical situation in which they compete should have an influence on our judgment, and he seems to believe that our ingenuity will *and should be permitted to* react differently to different types of influence: 'The human faculties of perception, judgment, discriminative feeling, mental activity, and even moral [and, so we might add, methodological] preference are exercised . . . in making a choice' and they are 'improved' by such an exercise.[3] Should the improved facilities not be given the chance to make improved judgments on the basis of improved standards?

5 This reasonable and rather sophisticated question which makes methodology part of the historical process did not survive. It certainly

---

[1] *Op. cit.*, p. 20.

[2] 'Mill' always means the Mill of *On Liberty*—not the Mill of the *Logic*: Mill was one of those rare individuals who can be persuaded by argument to completely change their point of view.

[3] Popper [1972], p. 252.

was not raised by the logical positivists, those pugnacious champions of clean living in the sciences. Quite the contrary, they implicitly denied that the development of methodology had anything to do with the development of human talents. They set up rules of procedure that treat all theories in the same manner, independently of the historical situation in which they are judged, independently of any development in our methodological imagination that might take place. History which had started playing a role in the evaluation of theories is now again ignored. Compared with Mill (and Hegel) this is a tremendous step back and, mind you, this step is not the result of arguments, it is the result of an uninformed allegiance to earlier non-historical methodologies. Karl Popper shares the basic ideology of logical positivism. This is why the story can from now on be told in at least two different ways.

The one way is the way in which Imre Lakatos has told it, in his essay 'Falsification and the Methodology of Scientific Research Programmes'.[1] Lakatos relates his ideas to the ideas of Karl Popper. He accepts the general fallibilistic framework of Popper's philosophy. The changes he introduces are: (*a*) standards are separated from rules of acceptance and rules of rejection; (*b*) they are not applied to individual theories, but to sequences of theories which are connected by certain techniques of modification, so-called research programmes; (*c*) the elements constituting a research programme need not be rewritten, or 'reconstructed' in accordance with the rules of a particular logical system, they can be used in the form in which they occur in science, leaving unchanged all the logical and empirical faults they possess; in this way both theories and the forms of life of which they are part can be examined; (*d*) the standards judge the evolution of a research programme over a period of time, not its shape at a particular time; and they judge this evolution in comparison with the evolution of rivals, not by itself; it is therefore possible for a research programme to remain within science even if all the theories it produces are self inconsistent and beset by oceans of anomalies; a research programme is called 'progressive' if it makes predictions that are confirmed by subsequent research and thus leads to the discovery of new facts. It is called 'degenerating' if it makes no such predictions but is reduced to absorbing the material discovered with the help of the rivals. It is important to repeat that the theories which constitute progressive research programmes may be full of anomalies, they may contain hidden assumptions, they may be imperfectly formulated—all this does not matter; if the sequence of theories produced by the techniques of the research programme, by its so-called heuristic leads to new discoveries, then the whole sequence leads to progress, and that alone is decisive (this is why Copernicanism can be said to be progressing over Ptolemy despite the fundamental dynamical and optical difficulties of the theory); (*e*) the standards judge research programmes, they do not tell the scientist what to do. This feature of the methodology considerably reduces the Puritanism and self-righteousness of a scientific debate. There is no rule that tells a scientist to remove a degenerating programme—and rightly so, for a degenerating programme may recover, and come out on top. (Such developments occurred in the case of atomism, the temporal finitude of the world, the moving earth. All these programmes advanced and degenerated numerous times, and all of them are now solid parts of science.) It thus may be *rational* to pursue a research programme on its degenerating branch even after it has been overtaken by its rival. But it is also rational to build *institutions* which inhibit the pursuit of degenerating programmes unless they show some promise in the form of new predictions. In this field everyone can choose freely and rationally. Imre Lakatos chooses the second way and gets much propagandistic mileage out of calling the second way and the further restrictions it implies also 'rational'; (*f*) the standards can be evaluated in a way that is very similar to the evaluation of empirical research programmes: we regard them as *historical* research programmes and the 'predictions' that decide *their* progressiveness and *their* degree of degeneration are predictions about basic *value* judgments (such as 'Einstein's theory of general relativity of 1915 is better than Newton's celestial mechanics in the shape in which it was available at the same time').

---

[1] Lakatos [1970].

Comparing these points with Popper we see that some elements have been added (such as the procedure for the criticism of standards), overly strict rules, such as the rule of falsification have been removed and the discussion has been brought much closer to scientific reality than it was in the *Logic of Scientific Discovery* (research programmes instead of theories; scientific rather than logical formulation of research programmes however great the 'faults'; standards judge developments rather than the situation at a particular time; and so on). Comparing them with Mill, we notice a very different kind of change. Mill provides an open framework that mirrors some decisive features of science and that can be supplemented with the results of concrete historical research. Nowhere is it necessary to correct Mill, or to soften some formulations. The route from Mill to Popper to Lakatos is a circuitous route, it involves a step back from scientific reality whereas the route from Mill to Lakatos is a direct route, one comes closer and closer to science. It is therefore more natural, from the point of view of ideas (from the point of view of the third world, as Lakatos would say) to see Lakatos as a development of Mill than to see him as a development of Popper. And this is the standpoint I have adopted. The reason why Popper is a step back is, of course, that he has been influenced by the philosophy of logical empiricism.

6 Mill and the Darwinists evaluate a theory—and that for them is something very similar to what Lakatos calls a research programme—by comparing it with the historical background in which it is embedded. The logical positivists evaluate a theory—and that is now a *statement* expressed in a language that differs decisively from the language of most scientists—by comparing it with some *elements* of this background, so-called observations which, moreover, have been separated from history and have been made to conform to some abstract demands. Logical Positivists restrict their framework of research, they make it more narrow, they move away from actual scientific practice. The historical understanding of the sciences that had advanced so far with Mach and Duhem diminishes and becomes almost non-existent (in the enlightenment the situation was pretty much the same). But modern science, and for that matter any kind of science, is too complex and too problematic to be aided by the simple models of the philosophers. So the attempt to *actively participate in its development*, to anticipate future stages, to contribute to the solution of pressing problems which played such a large role in each is gradually replaced by the more traditional attempt to *understand what exists*. But understanding now no longer means a relation to the sciences that enables the person so related to make discoveries. It means dealing with the problems that arise when a special set of abstractions is chosen, so that it is now this set and not science that is being discussed: instead of a scientific philosophy we have another school-philosophy that borrows the *name* from science but has almost nothing to do with it. Popper showed how a model formed from this set can be made to work smoothly, *i.e.* without violating demands of logic together with another demand which seemed to come from science, *viz.* the demand of *non ad-hocness*. But, of course, the fact that the model runs smoothly does not mean that one has understood science, for *the model and science have little in common.* Mill, on the other hand, is much closer to science for although his theory is very general it deals with events and elements *that actually do occur in science* and he tells us, correctly, though in a general way, how they are arranged. He also provides for a criticism of standards though he does not specify the particular way in which the criticism is to be carried out. He specifies some general features of the criticism, the presence of a multiplicity of theories and facts rather than of one theory and of facts 'relevant' to it. He also allows us to evaluate elements of science that

cannot be neatly written down on paper. We are on the right way, all we need are more details. Imre Lakatos gives us the details and brings Mill closer to science. He also corrects Mill, slightly, for while Mill had excluded mathematics from his account, Lakatos shows that informal mathematics, which is that part of mathematics that is responsible for its growth, can be incorporated into the fallibilistic scheme. He removes the last Aristotelian element; the element of necessity, from modern science.[1] And he corrects the idea, originated by Xenophanes and Parmenides that knowledge deals with unchanging entities and must be *presented* in the form of principles describing such entities. He reverses a trend that has dominated western thought for more than two millennia. What can we say about his account?

7 It is definitely preferable to Newtonian accounts up to, and including those introduced by logical empiricism. It leaves it to the scientist to formulate his theories and to invent ways of criticising and of improving them. It assumes that the scientist, being in touch with the concrete reality of research will have developed his talents to a sufficient extent to know how to deal with the problems that arise and how to change his theories to solve them. On this point Lakatos agrees with Mill's expectation that our faculties of judgment are improved by being exercised in making a choice. He also agrees with the idea that there is something like an instinct for right and wrong which is developed by the concrete interaction with the historical material and which cannot be fully articulated. Even 'external' demands, such as the demands of logic, or the demands of theology may influence the activity of the scientist but neither they, nor instinct are evaluated *by themselves*, it is not *automatically* assumed that a science that is run by instinct, that is logically acceptable and theologically safe is better than a science that does not agree with such conditions. Whatever influences science is *evaluated*. It is evaluated by the amount of progress it brings about, in accordance with the standards.

The standards themselves are defended historically, as guiding modern science *at its best*. This being the case they can be directed against developments in science itself as not agreeing with what is best in it. Already this possibility shows that basing the standards on history is not a circular procedure. On the other hand we see now also that we cannot turn the standards against magic, or against Aristotelianism for all we can say is that Aristotelianism is *different* from what is best in modern science, not that it is worse. Here Mill's point of view is definitely preferable to that of Lakatos. Mill chooses his method not because it fits a certain subject. He chooses it because he believes that acting in accordance with it will improve the individuality of the man so acting, will make him a better thinker and a happier man.

---

[1] *Cf.* Lakatos [1963–4]. In a way Lakatos here seems to arrive at a position similar to Quine's. It is similar to Quine in the same way in which Newton's account of his law of gravitation is similar to Hooke's.

There is one feature of the methodology of research programmes that distinguishes it decisively from all other methodologies and that is: to apply it we must study history in considerable detail. A naive falsificationist who wants to judge a particular theory will either study the theory itself (or, in most cases, a distorted image of it in some pidgin logic, a so-called 'reconstruction' of the theory), or he will study the relation of the theory to 'the evidence'. Once he has found a particular piece of evidence, *no matter in what historical surroundings*, that contradicts the theory, the theory must go. The methodology of research programmes does not examine theories, it examines research programmes which are sequences of theories connected by certain hard cores, heuristics and intuitive attitudes *not all of which need to be formulated explicitly*. Already at this point research must go much deeper than the research of a Popperian ever would. Secondly, the methodology of research programmes does not only examine research programmes by themselves, it also examines them in comparison with rivals. So the investigation spreads and must reach, ultimately, every research programme that competed at the period in question. In the sixteenth and seventeenth centuries this includes research programmes which we today would regard as non-scientific (theological, *etc.*). Thirdly, we must examine whether the predictions made were novel predictions rather than repetitions of what was already known. This means, we must also know something about the expectations of the age and about the 'accepted facts'. To be able to make a judgment we must know a large amount of material that belongs to the history of ideas and that is often absent even from the customary *historical* accounts. All this taken together leads to the judgment that the research programme progressed, or degenerated *in a certain period*, and that its heuristic was barren, or fruitful, *in this period*. Nothing else can be said. The research programme may have been excellent before the degeneration set in, and it may recover immediately afterwards. It is evident that standards such as these will throw an entirely different light on subjects such as *astrology* which up till now have been condemned either out of sheer prejudice, or on the basis of naive criteria. It may well turn out that astrology, despite all its internal contradictions and despite all the known anomalies, progressed in the fifteenth century, and degenerated afterwards, and has not recovered since. It is quite possible that we shall make this discovery. But to make it, we need much more material than is known about astrology today, when the general belief is that the case has been settled long ago, by naive falsificationist standards.

8 So far I have compared the theory of research programmes with theories of science which assume that scientific achievements can be evaluated, that some achievements are better than others and which also assume that the standards of evaluation can be openly discussed and criticised. Knowledge, for such theories, is an objective process to which every generation makes its contribution, it is constantly improved and modified. This view of knowledge was first stated with brilliant clarity by Aristotle and it explains why Aristotle starts the discussion of every problem with the ideas of his predecessors. It has emerged that Lakatos gives by far the best account of such an objective knowledge yet in existence. The account is not free from fault, it can be improved, but so far it is the best. But there are other theories of science which drop one or the other of the assumptions of the objectivists. *Skepticism*, for example, assumes that theories and world views are all equally good or equally bad and that any attempt to distribute praise and blame is sooner or later bound to run into difficulties. Another school, called epistemological *Stalinism* by Lakatos (in some of his talks, not in any one of his publications) assumes that the evaluation of theories depends on the judgment of some Great Man or of some Great Group: good theories are those theories which great scientists, or groups of great scientists *say* are good. Science, according to this view, is a complex adaptive mechanism run by complex beings, scientists, who often do not know what they are doing but who have developed a kind of instinct, a *Fingerspitzengefühl* that enables them, singly, or collectively, to do the 'right thing' and who therefore succeed in the end. The reply to the Stalinists is simple. The methodology of research programmes does not deny that scientists often use procedures of which they cannot give an explicit account, nor does it want to eliminate such procedures from science, as we have seen. But it asserts that there is a way of evaluating their *effects*, with the help of the standards, and it adds that this is the way in which science has been run, consciously or unconsciously, in the past. To remove *this* assertion, one has to criticise the standards. So far no Stalinist has offered the required criticism. A different point, made by Kuhn, cannot be answered so briefly and it seems that it cannot be answered at all. Kuhn, according to Lakatos, has to use 'mob psychology' to explain revolutions, but the methodology of research programmes can give a 'rational' account. One does not see how this can be true. The revolution that introduced modern science also introduced the standards, and so it cannot be judged by them (remember that the only reason Imre Lakatos gives for the standards is that they underly what is best in modern science). Subsequent revolutions always involve various parties. Given the standards the actions of any party are as 'rational' as the actions of any other party. The fact that one party succeeds and the other fails is therefore, a purely sociological phenomenon, inexplicable on the basis of the standards—and this is what Kuhn tries to say. The Skeptic, on the other hand, will be utterly unimpressed by standards which are compatible with any action. For him such standards are standards in appearance only and he will claim that in actual fact one research programme is now admitted to be as good as any other research programme. It seems that Imre Lakatos was somewhat impressed by this point and towards the end of his life he tried to beat skepticism not by an appeal to argument, but to commonsense: where is the skeptic who walks out of the window instead of taking the elevator? But this reply will not do. First of all, because it is very doubtful whether this difference in behaviour has been brought about by a competition of research programmes, evaluated in accordance with standards of progressiveness: in some species reactions against the appearance of hight are innate. Secondly, no skeptic has ever held that skeptics are brave men and free of prejudice. The question is: are they *right*—and to *this* question the methodology of research programmes does not seem to give an answer.

9 We see that the methodology of research programmes, despite the tremendous progress made, is still full of faults and has not overtaken its rivals in all points. For a follower of Mill this is not surprising. Nor will such a fallibilist now turn to Kuhn or to the Skeptics. The reason is obvious. Kuhn gives a very *general* account of scientific (and other) developments while Lakatos forces the researcher to look for *specific details*: what is the hard core; what is the protective belt; what is the heuristic; how are these elements used in the explanation of a particular phenomenon; and so on. As an instrument for carrying out research in the history of ideas Lakatos's theory is vastly more sophisticated than Kuhn's, and so it will lead to more detailed research, and to more discoveries. The discoveries may in the end turn against Lakatos, but that does not discredit him *today*, when no other theory provides an equally detailed inventory of suggestions. This is also how the issue with skepticism may be resolved. Speaking in an *abstract manner* the skeptic seems to be right. But he may be wrong in *this* world, for this world may be open to inspection to such an extent that we can indeed separate better and

worse theories and give reasons for our preference. The material which the followers of Lakatos will discover in the course of their research is relevant to this question. It is difficult to see how the skeptic, following ***his own*** philosophy, could ever unearth it. He needs the methodology of research programmes to give him those facts that might show that the world is not as orderly as assumed, and that Skepticism is the correct position to take in it.

University of California, Berkeley

REFERENCES

BACON, F. [1608]: *Novum Organum.*
BOLTZMANN, L. [1905]: *Populaere Schriften*, Leipzig.
BORN, M. [1949]: *Natural Philosophy of Cause and Chance.*
COHEN, M. (*ed.*) [1961]: *The Philosophy of John Stuart Mill.*
DODDS, E. R. [1957]: *The Greeks and the Irrational.*
EINSTEIN, A. [1936]: 'Physics and Reality', *The Journal of the Franklin Institute*, **221**, pp. 313–82.
FEYERABEND, P. K. [1970]: 'Classical Empiricism', in R. E. Butts and J. W. Davis (*eds.*): *The Methodological Heritage of Newton.*
GINGERICH, O. [1974]: 'Aesthetics versus Crisis in Copernican Astronomy', in A. Beer (*ed.*): *Vistas in Astronomy*, **17**.
HEGEL, G. W. F. [1817]: *Encyclopaedia*, Lasson (*ed.*).
KOCHER, H. R. [1953]: *Science and Religion in Elizabethan England.*
LAKATOS, I. [1963–4]: 'Proofs and Refutations', *British Journal for the Philosophy of Science*, **14**, pp. 1–25, 120–39, 221–43, 296–342.
LAKATOS, I. [1968]: 'Changes in the Problem of Inductive Logic' in I. Lakatos (*ed.*): *The Problem of Inductive Logic*, pp. 315–419.
LAKATOS, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes' in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*, pp. 91–196.
LAKATOS, I. [1971]: 'History of Science and Its Rational Reconstructions', in R. C. Buck and R. S. Cohen (*eds.*), *Boston Studies in the Philosophy of Science*, **8**, pp. 91–135.
MACH, E. [1919]: *Die Leitgedanken meiner Naturwissenschaftlichen Erkenntnislehre und ihre Aufnahme durch die Zeitgenossen*, translated in S. Toulmin (*ed.*): *Theory and Reality.*
MILL, J. S. [1859]: *On Liberty.*
NIETZSCHE, F. W. [1873]: *Wahrheit und Lüge im Aussermoralischen Sinn.*
NILSSON, M. P. [1949]: *Greek Folk Religion.*
POPPER, K. R. [1972]: *Objective Knowledge.*
SZABÓ, A. [1969]: *Die Anfänge der Griechischen Mathematik.*
WILAMOWITZ-MOELLENDORF, U. von [1955]: *Der Glaube der Hellenen.*

Brit. J. Phil. Sci. 26 (1975), 19–26 *Printed in Great Britain*

The Problem of Atomism*

by HEINZ POST

Atomism has been defined and classified elsewhere[1] as the programme to explain all changes in terms of invariant units. I distinguish between material and abstract atomism. The former refers to 'atoms' that are, at least in principle, located in space; the latter does not, but introduces abstract units of explanation instead.[2] In the light of these definitions I hope to show that theories as diverse as those of Democritus, Leibniz and Newton and present elementary particle theories may be seen as attempts to deal with a common fundamental problem, the difficulty of reconciling a world of interacting parts with atomism, which ideally requires independence for its atoms.[3]

'The world is so full of a number of things'. It is the task of science to *explain* this variety, including variation in time, *i.e.* change.

Any real explanans must be in terms other than the explanandum. So, if we want to explain variety, the multiplicity of things, we must explain it in terms of non-variety, of unity. We explain the many in terms of the one, the unit. Similarly, change must be explained in terms of the unchanging, the invariant.

There are two extreme ways of dealing with the problem of change: we may deny change as an illusion (Parmenides), or we may embrace change in the spirit of 'everything flows' (Heraclitus), or, worse, 'anything goes' (Anaxagoras).[4] The latter view despairs of explanation; the former despairs of reference to empirical data.

We shall be concerned with the third way as exemplified by Democritus, the 'search for the pole at rest in the flight of events'. In this programme of reducing multiplicity to unity, we introduce *units* of limited variety, preferably all alike. We are looking for invariant units more or less hidden behind the world of change. The atomistic programme is to explain everything, but everything, in terms of a denumerable number of identical invariant units, or at least units of limited variety, *i.e.* of a small number of species. Ideally, we would want one species only. By the same token of simplicity we make our units as large as possible (compared to experimental accuracy) so as to make the possible combinations as restricted as possible. At any rate, we make our units *finite*.[1] If the variety of appearances in the universe is infinite, we would want in the material case an infinity of arrangements of individual units in space (relative location being the sole primary quality) to account for the infinity of appearances.

---

*The substance of this article was presented to the British Society for the Philosophy of Science in a Presidential Address in April 1972.

[1] Post [1968]. These classifications include the distinction between 'intermediate' and ultimate atoms: indivisibility, quite apart from philosophic difficulties, is not, in my view, the fundamental criterion for the explanatory role of atomism. The chemical atom is a very good explanatory unit at an intermediate level.

[2] Democritus's atoms are examples of the former; linearly additive quantum numbers such as isobaric spins are examples of the latter.

[3] There are, of course, other problems, traditionally associated with atomism, such as (1) the problem of the void, and (2) the problem of perfect hardness (see *e.g.* Scott [1970]); but these are, in my view, less fundamental.

[4] Although Anaxagoras did maintain a principle of conservation (Fragment 17), he did not cash in on this principle scientifically; that is to say, he did not draw consequences testable by independent criteria. In fact he specifically denied this possibility (Fragment 7). He was a pluralist.

In the abstract case, we would want the state to be characterised by few, ideally only one, parameter, and would interpret the variety of appearances as a superposition of states. Aristotle's theory of elements-in-potentia (wet-dry; hot-cold) is an example of this. (We have an abstract double dichotomy similar to Aristotle's in present nucleon theory, in terms of 'ordinary spin' and 'isobaric spin'.)

The tendency in recent physics has been towards descriptions no longer in terms of individuals, *i.e.*, objects localizable in space-time, but in terms of states (filled or otherwise). We shall find that this shift also overcomes some logical difficulties. This shift corresponds to a shift from material atomism to abstract atomism. In abstract atomism, we may retain the notion of space, or we may not. Even when we retain it, we no longer allow localisation of particles, we no longer have individuals.[2] We speak instead of states being realised or filled. In the case of a Fermi field, a state either is occupied or it is not. This exhausts the possibilities. In the case of a Boson field, a state may be 'occupied' by any number, denoting so-called 'particles'.

In both material and abstract atomism we search for invariant elements. Of course, in a sense this search is common to all theories, atomistic or otherwise. As has been well said: 'The irrelevance of the values of some parameters to the invariant correlation of some other parameters is *not* a matter of convention'. The power of the usual type of 'if-then' law is precisely measured by the irrelevance range of some parameters. These 'if-then' laws (temporal, experimental) are, after all, nothing but generalisations of Moliere-type 'explanations', and very good laws they are. We do not just say this powder puts you to sleep because it is soporific, we say all powders obtained by precipitation from the juice of poppy seeds put you to sleep because all powders obtained by precipitation from the juice of poppy seeds are soporifics, whether the poppy be red, yellow or white. Atomism goes deeper than that: it even essays to explain the 'irrelevant' data, the colours of the poppies, at a deeper level.

---

[1] Indeed, Boltzmann, who has been called 'Atomist bis ins Unmögliche' rejected even the mathematical infinitesimal calculus.

[2] Non-localizability is necessary, but not sufficient, for non-individuality. (See Post [1963].)

Is atomism, then, coextensive with explanation in general? No: in the material case we have the counter-example of a reconstructed Cartesian continuum theory; in the abstract case we have the holistic Gestalt theory. Nevertheless, counter-examples to atomism are few and far between, and atomism certainly exhibits the typical virtues of a good theory, the search for invariance.

This requirement of invariance, however, raises an apparently fatal difficulty: these invariant factors must be not only independent of external factors, but must be independent of each other, otherwise they would not be, in fact, invariant. Interaction between factors would imply change in the factors.[1] We are looking, then, for invariant factors of explanation. Here I agree with J. S. Mill rather than with his godless godson. '. . . I shall give the name of the Composition of Causes to the principle which is exemplified in all cases in which the joint effect of several causes is identical with the sum of their separate effects.' (Mill [1872], Bk. III, Ch. VI, 'Independent and Emergent Qualities'.)

The difficulty is that the whole interest in science seems to arise from the non-independence of the postulated units. A theory of material atoms that does not introduce interaction between them is either ludicrously false[2] or ludicrously weak. The same applies to a theory of abstract atomism that does not introduce correlation between the predicates. The universe is more than an assembly of independent atoms, just as 'a house is more than a pile of bricks'. The real rival of atomism is holism.[3]

[1] Thus, in present theory the 'chemical atom' is not the explanatory unit to explain chemical change. Chemical atoms are severely changed by interaction. The age-old philosophical problem of the 'persistence of elements' in compounds is explained as the recoverability of systems of protons and electrons in well-defined quantum states. This in turn raises the problem of explaining Coulomb interaction in terms of a (photon) field theory, and so on.

[2] A variant of dealing with this problem is exhibited in the history of early models of the chemical atom by Bohr's extraordinary *ad hoc* 'Principle of Invariance of Quantum Numbers', Bohr [1923]) which admits interaction but postulates (falsely, because inconsistently) that individual quantum numbers are maintained.

[3] 'Instead of starting with the elements and deriving the properties of the wholes from them a reverse process is necessary, *i.e.*, to try to understand the properties of parts from the properties of wholes. The chief content of Gestalt as a category is this view of the relation of parts and wholes involving the recognition of intrinsic real dynamic whole-properties' (Koffka, as quoted in Nagel [1961], p. 391). Compare also Lesniewski's merology starting from the notion of parts. In different ways, both Bohr and some philosophically-minded biologists appeal to some forms of holism.

P869

The fact that in quantum mechanics we have an elementary particle associated with every interaction is but a reflection of this difficulty: nucleons interact via mesons, non-trivial meson theories are non-linear and require interaction between mesons associated in turn with other particles.[1] Big fleas have little fleas, and so on *ad infinitum*. The programme of atomism requires a regularity principle leading to a lowest level of ultimate atomism. A theory postulating that a specified set of elementary particles is, in fact, ultimate could be a perfectly respectable scientific explanatory theory. It is by no means the case that a theory postulating little fleas necessarily constitutes a theory that is deeper in the sense of being better. On the other hand, a theory reducing all elementary particles at one level to a continuum at a deeper level may indeed be an advance.

Thus I entirely disagree with Russell, who says 'In this world, whatever is complex is composed of related simple things, and analysis is no longer confronted at every step by an endless regress'. (Russell's Doctrine of External Relations, 'The Nature of Truth', quoted in his [1959], p. 61.) I do not understand the point of Russell calling his theory logical atomism,[2] or, indeed, the point of his theory at all, since it allows such 'atomic' statements as $R(x_1, \ldots x_n) = 0$ involving n-tuple relations, where n for all we know may even be non-denumerably infinite. If things 'compose' complexes, they are not simple. This kind of analysis *is* confronted at every step by an infinite regress. Unlike Nagel, I agree with Mill in drawing a distinction between linear and non-linear composition. The former is used in the atomistic programme of synthesis; the latter is not. A categorical distinction *is* possible, in that the former allows any of a non-denumerable number of superpositions as a *possible* solution, whilst the latter does not.

One of the ways, then, of overcoming the fundamental difficulty of atomism, to wit that we have to introduce interaction in the material case, or correlation in the abstract case, to obtain more than a trivial universe, is to draw a demarcation between linear and non-linear dependence. Linear superposition theories are to be allowed, non-linear theories are not. Mill's composition of causes is precisely a case of linear superposition. If we have a metric we must introduce linearity to make the point about perfect atomism. But even in the absence of a metric we may point out the invariant correlation of certain qualities on a yes-no basis: we may for instance observe

[1] Any 'ultimate' atoms in quantum mechanics necessarily interact with each other in so far as they couple to any (other) species of particles.

[2] Significantly, Russell does describe himself as a pluralist; but in spite of his disclaimer, I consider his (mistaken) ground for calling himself an atomist to lie in his view of natural science and not of logic. It is a naive view of the undeniability of sense data. He is a positivist. The only excuse for calling his theory atomistic lies in the postulated *logical* independence of his *sentences*. His 'simples' are far from independent. Not every 'analyst' is an atomist.

a hundred per cent correlation between the possession of blue eyes and fair hair. Even such total correlation (let alone any weaker one) may well be hidden when starting at a naive observational level. It is analysis that reveals the, at least linearly, independent factors. It is precisely linear dependence that is removed by diagonalisation. *Provided* there is no other dependence, we have thus reached the true (abstract) atoms.

In general, we are faced with the situation of *partial atomism* in present-day theories. We explain ordinary if-then laws at a deeper level by atoms, but these atoms in turn have to be supplemented by other laws. These laws are, admittedly, more general, simpler than the original laws at higher level, but they are not, in general, just expressions of linear superposition. They are frequently non-linear interaction laws in the material case, or non-linear dependences of parameters in the abstract case. If we only have atoms, without laws, our theories are in danger of being empty. If we have only laws, without atoms, we may be unduly shallow.[1]

Partially atomic theories are ultimately unsatisfactory. There are ways of overcoming the difficulty other than an appeal to a distinction between internal and external relations, a distinction which seems to me to beg the question. Is the life-time of a neutron an internal or external parameter? (It is certainly changed by interaction with a proton.) One radical philosophic way is, in effect, to replace the whole of dynamics by mere kinematics to let atoms move according to pre-established harmony. Another way of meeting the difficulty of interaction is to introduce randomness, a notion that has haunted atomism throughout its history. The link between the ultimate atoms and the higher level may have to be probabilistic. We have a model of ultimate independent atoms explaining a law-obeying behaviour at a higher (macroscopic) level in the relation between molecules moving at random and Boyle's law by way of a probabilistic law of large numbers.[2] Again, Einstein's relatively shallow theory of photons does not require interaction between photons to explain the black-body law. Boltzmann had to point out to Planck, the absolutist insisting on numerical atomism, that he could only obtain his law by the introduction of randomness (natural radiation).

But we do have, finally some cases of *perfect* atomism, of absolutely independent atoms, though usually in the abstract case.[1] We have the analysis of motion into normal modes,[2] and the analysis of correlation matrices into principal components.[3] We can certainly find components that are at *least* linearly independent. The deeper atomistic programme is to 'diagonalise' the whole universe into completely independent factors.

---

[1] Laws may perhaps be reduced entirely to symmetry requirements (Wigner) and hence reduced to abstract atomism, the invariants, and these invariants may be truly independent: we may have linear addition of equal spin units, regardless of other parameters. (As against such unexplained correlations as the correlation of electronic mass and charge.)

[2] In this, as in other respects, we find analogous histories of atomism in the physical and the biological sciences. Already in 1908 Exner suggested that all deterministic laws are basically large-number versions of probabilistic laws.

Perfect atomism may be exemplified eventually by a possible elementary particle theory. Elementary particle theory has to be judged by the standards we have used before. We are not recommending the invoking of group theory for the purpose of empty numerology. What I do wish to point out is that in an elementary particle theory with a suitable interpretation, a conservation principle or the corresponding symmetry principle may be sufficient at an ultimate level to draw independently testable consequences at a higher level.[4] This would correspond to Wigner's programme of reducing all laws to super-laws in the form of invariance principles.

At present holism *appears* to win if we match it against a naive material atomism of elementary particles: neither charge nor mass, either gravitational or inertial, is definable for a particle in isolation, even in classical physics. It is only the *interaction* which is defined operationally; it may be represented artificially as the product of 'charges' associated with the separate particles. Again, the operational intertial mass (the reduced mass) of a particle only tends to its 'own' mass in the limit of interaction with an infinity of similar particles. The 'universal constants' describing a particle are global, and presumably are what they are because the universe is as it is cosmologically. 'Renormalisation' in quantum mechanics, on the other hand, *starts* from the unrealistic 'naked' (independent) atom and has to add the effect of interaction.

Quantum mechanics *might* provide us with a perfect abstract atomism (spins independent of all other parameters, conserved, and linearly additive) but at present we have to pay the price of abandoning realism if we buy quantum mechanics.[5] Are atoms real? I do not really understand the meaning of the question. Perhaps reality should be measured by the range of invariance. I think I should be classified as an atomist who does not believe in reality, in the sense that I do not believe in things, objects, as primary. I am not a language-based noun-ontologist, and I can hardly wait for a logic without individuals.

---

[1] An example of perfect material atomism is Hauy's theory of crystal structure. Here the geometric formulation is taken as ultimate. Historically crystallography only began to flourish when x-ray analysis both allowed and forced a deeper structural explanation. Unfortunately, this deeper structure in turn called for an underlying mechanism. This programme has not yet been fulfilled, partly owing to a 'computation gap'.

[2] See *e.g.* Webster [1959], p. 163.

[3] See *e.g.* Lawley and Maxwell [1963].

[4] Cf. Braithwaite's '3-factor theory' in his [1964], pp. 56–68.

[5] According to D'Espagnat ([1973]) recent experimental results on correlation between distant measurements force us into holism anyway. Bohr has generally been interpreted to be a holist.

One final remark: the programme of synthesis seems to me at least as important as analysis.[1] The categorical importance attached conventionally in textbooks and elsewhere to the experimental 'production', 'detection', or 'isolation' of the elementary particle, the atom, such as the 'detection' of the free neutrino seems to me exaggerated.[2] The distinction between evidence derived from such experiments and the vast body of confirmatory evidence involving the particle in other ways by way of synthesis is hard to maintain or justify,[3] except on grounds of expense. The building of White Oliphants may, of course, be an end in itself for some people. But I think there is some point in the following anecdote: When the 'discoverers' of the neutrino at last succeeded after years of effort to obtain some additional counts attributed to inverse beta-decay, they telegraphed Pauli, saying: 'The neutrino exists'. Pauli replied: 'I know'.

There are two persistent heresies in the history of atomism: there is 'Anaxagoras's heresy': the unnecessary loading of properties on to the atom, the failure to appreciate the 'emergence' of secondary properties; the other heresy is the 'Mach heresy' of insisting on 'seeing' atoms in isolation.

Democritus had the true atomic programme, but his material atomism did not get off the ground because he had no interaction. His vortices and like-to-like effect were lapses that did not work. Newton adulterated material

---

[1] In some respects, Atomism has remained only a programme; for example the splendid and plausible attempt to explain the macroscopic geometric shape of crystals by analysis into atomic lattices has not resulted in a successful theory synthesising crystals from atoms.

[2] 'While careful reasoning from experimental evidence gathered about all terms in the beta-decay process—except the neutrino—may support the inference that a neutrino exists, its reality can only be demonstrated conclusively by a direct observation of the neutrino itself. If the neutrino is a real particle carrying the missing energy and momentum from the site of the beta decay, then the discovery of these missing items at some other place would demonstrate its reality. Thus if negative beta decays as in equation (1) could be associated at another location with the inverse reaction:

$$\nu_{-}+p^{+}\rightarrow\beta^{+}+{}^{0}n \tag{6}$$

which is observed to occur at the predicted rate, the case would be closed .... Completion of the term-by-term checks of equation (6) thus demonstrated that the free neutrino is observable in the near vicinity of a high-power fission reactor' (Cowan and Reines [1956]).

[3] It is not *necessary* for quarks to be isolated in the free state for them to be real. 'Of course, the whole quark idea is ill founded. So far quarks have escaped detection. This fact could simply mean that they are extremely massive and therefore difficult to produce, but it could also be an indication that quarks cannot exist as individual particles but, like phonons inside a crystal, can have meaning only inside the hadrons' (Kokkedee [1969], p. ix). Mendeleev [1876] on the question of 'isolation' of molecules: '. . . the chemists . . . know that they will not get to extract and see these small molecules in the way that a vegetable cell is extracted'.

atomism with forces.[1] Quantum mechanics has abstract atoms but no realism. The difficulty in its 'theory of measurement' is again basically to be traced to the impossibility of reconciling linearity with interaction with the apparatus.

The conclusion of this paper is that perfect atomism is possible in principle: we can get restrictive laws by suitable (higher-level) interpretation in 'observational' terms of atoms independent at an ultimate (deepest) level.

University of London.

REFERENCES

BOHR, N. [1923]: 'Linienspektren und Atombau', *Annalen der Physik*, **71**, pp. 228–88.

BRAITHWAITE, R. B. [1964]: *Scientific Explanation.*

COWAN, C. L. and REINES, F. [1956]: 'The Neutrino', *Nature*, **178**, pp. 446–9.

D'ESPAGNET, B. [1973]: 'Quantum Logic and Non-Separability' in J. Mehra (*ed.*): *The Physicists Conception of Nature*, pp. 714–36, Dortrecht: D. Reidel.

KOKKEDEE, J. J. J. [1969]: *The Quark Model.*

LAWLEY, D. N. and MAXWELL, A. E. [1963]: *Factor Analysis as a Statistical Method.* London: Butterworth.

MCGUIRE, J. E. and RATTANSI, P. M. [1966]: 'Newton and the Pipes of Pan', *Notes and Records of the Royal Society of London*, **21**, no. 2, December.

MENDELEEV, D. I. [1876]: *Materials for a Judgement on Spiritualism*, translated by J. R. Smith. Russian original in D. I. Mendeleev, *Collected Works*, **24**, pp. 194–9, Academy of Sciences of the USSR, 1954.

MILL, J. S. [1872]: *A System of Logic.*

NAGEL, E. [1961]: *The Structure of Science.*

POST, H. R. [1963]: 'Individuality and Physics', *The Listener*, 10 October, pp. 534–7.

POST, H. R. [1968]: 'Atomism 1900', *Physics Education*, 3, pp. 225–232.

RUSSELL, B. A. W. [1959]: *My Philosophical Development.*

SCOTT, W. L. [1970]: *The Conflict between Atomism and Conservation Theory 1644 to 1860.* Macdonald & Co.

WEBSTER, A. C. [1959]: *The Dynamics of Particles.* Dover.

[1] 'For the whole burden of philosophy seems to consist in this—from the phenomena of motions to investigate the forces of nature'. Newton's anti-hylozoistic notion of force is in the neo-platonist tradition with God not yet separated from His universe (McGuire and Rattansi [1966]).

Brit. J. Phil. Sci. 26 (1975), 27–39 *Printed in Great Britain*

Doubts about the Correlation Thesis*

by ROBERT C. SOLOMON

In the recent dispute over the psychophysical identity theory (*IT*), there has been apparently unquestioned agreement upon the plausibility of a second thesis which I shall refer to, following many authors on the topic, as the 'correlation thesis' (*CT*). The *CT* is generally recognised as a well-confirmed or at least highly plausible[1] *factual* hypothesis to the effect that there is a uniform and demonstrable correlation between brain processes of certain kinds and sensations. With the acceptance of this empirical thesis, the mind-body dispute concerns the ontological and theoretical commitments which ought to be imposed upon the *CT*. The *IT* is a currently fashionable ontological interpretation of the *CT*, often defended with an eye to maintaining the 'unity of science' or materialism, sometimes defended on the weaker appeal to theoretical simplicity and parsimony. Against the *IT*, it has been argued (1) that the *IT*—as opposed to a 'mere' *CT*, is ultimately unintelligible (*e.g.* by Baier [1962]) and (2) that the identity theory yields no significant heuristic advance over and above the *CT* (*e.g.* by Brandt [1960]). But in each case, the *CT* is taken as the point of agreement and the debate focuses upon those theoretical and ontological issues which the *IT* adds to the *CT*. The *CT* is thought to be an empirical, non-theoretical and non-ontologically committed description of the data upon which the *IT* builds a non-empirical (*vis-a-vis* alternative theories, causal interactionism, parallelisms, epiphenomenolalism) and ontologically committed theory. In this essay, I want to argue that examination of the *CT* yields a very different picture of its proper rôle in these philosophical disputes, one which undermines the presuppositions of much of the current debate.

Received 21 *January*, 1974

---

* I should like to thank Bernard Gendron for his discussion and criticism and the referees of this *Journal* for their helpful comments.

[1] The *IT* is typically stated with a modest reference to future scientific investigations. J. J. C. Smart, for example, argues only that there are *now* no philosophical considerations that imply that science will not in the future show that sensations and brain processes are identical. If there is no *empirical* difference between *IT* and *CT*, I presume that the same modesty would insist that *CT* is a matter of future confirmation. My argument, then, is that there are now philosophical considerations which show that *CT* (upon which current versions of *IT* are based) is unacceptable. My argument ultimately has nothing to do with versions of *IT* in its most powerful form, *e.g.* Rorty's 'eliminative materialism', which is not based upon *CT* in any form.

1 What is the correlation thesis?

In its most general form,

CT: $(x)\,(\psi)\,(\exists\phi)\,[x \text{ is } \psi \text{ at } t \supset (\exists y)\,(y \text{ is } \phi \text{ at } t)]$

where 'ψ' is a variable for psychological states and 'ϕ' for neurological states, or simply,

$(x)\,[(x \text{ is an experience of type } \psi) \supset (\exists y)\,(y \text{ is a physiological state of type } \phi)]$

The converse of this thesis,

CT:* $(y)\,[(y \text{ is a physiological state of type } \phi) \supset (\exists x)\,(x \text{ is an experience of type } \psi)]$

and also the conjunction of *CT* and *CT** as a material biconditional might then be maintained on the basis of similar evidence, but it is currently of far less interest philosophically. What are primarily at stake are the most general physicalistic doctrines, notably the unity of science and materialism. And so there is far more concern with eliminating or reducing sensation-talk than there is with defending the thesis that every brain process of a certain type is correlated with some experience or other. Moreover, since '$a = b$' allows one to say with equal plausibility that 'a is nothing but b' and 'b is nothing but a', a true psychophysical identity thesis would support an idealist ('certain brain processes are nothing but experiences') as well as the materialist ('sensations are nothing but brain processes'). The idealist argument is not easily found today (although I have heard Charles Hartshorne argue a version of it), but it evidently was quite widespread in the late nineteenth century (*cf.* Nietzsche's argument against it in *Beyond Good and Evil,* Chapter 1, Section 20). Accordingly, I shall concern myself only with *CT.*

The conditional of *CT* is, as stated, material implication. Strict or 'logical' implication is surely too strong, for the correlation here is an empirical correlation, not a translation or entailment of psychological descriptions and physiological descriptions. But *CT* as stated is not yet a law.

Now any psycho-neurologist who would find the *CT* worth defending would not be content with this weak implication, but, as soon as he attempts to strengthen it, he finds himself in the midst of the *IT* controversy. Should the scientific status of the *CT* be that of a causal law?—or that of an inter-theoretic bridge principle?—or should it be modestly left as a 'mere' *CT,* a straight-forward empirical hypothesis which makes no attempt whatever to go beyond the inference from previous correlations to probable future ones? But these issues, as we find them discussed in the *IT* debate, are strictly non-empirical, and thus are part of the attempted move from *CT* (and *CT**) to *IT* (or its sister theories) and not part of the

CT itself. Accordingly, we shall leave our formulation of the *CT* in its weak material conditional form. The nature of the 'connection' between the antecedent and consequent events, between sensations and brain processes, is not part of the *CT*.[1]

We might note, innocently for the moment, that there is some discomfort in our quantification over 'experience'. It is not clear—either in our formulation or in the *IT* literature—whether the variable in question should be 'experience', 'sensations', 'experience of sensations', or something else. While the nature of such sensation talk has been extensively discussed in related philosophical areas, the conclusions have rarely carried over into the *IT* controversy.

The above form of the *CT* allows for a number of interpretations. Most importantly, there is a familiar ambiguity in the existential quantification over 'brain processes' between 'there exists some brain process or other . . .' and 'there exists a particular brain process . . .'. The first interpretation yields what I shall call the *weak* version of the *CT:* 'For every experience, there is some brain process or other'. The second interpretation is a stronger, (perhaps 'scientific') version of the *CT:* 'For every experience, there is a (more-or-less) specifiable brain process'. In the current *IT* debate, the modest appeal to 'future psycho-neurological investigations' might more accurately be reserved for the strong version only. While we are still admittedly ignorant about the precise nature of brain processes and their localisation in the central nervous system, we are even arrogantly confident of the general correlation of every experience with some brain process or other. And this, of course, is true not only of avowed materialists and scientific realists. Even the staunchest Cartesian would accept the weaker *CT* thesis. This confidence in the weaker version of the *CT* is not simply a matter of massive evidence; it is part and parcel of our modern version of the 'principle of sufficient reason'—without the more extreme commitments to materialism and scientific realism that sometimes accompany this scientific confidence. (The staunch Cartesian might also accept the stronger version of the *CT*, but must continue to attack the *IT*.)

The stronger version of the *CT* requires much more of the neurologist. It is not sufficient that he can register *EEG* traces during all waking and semi-conscious hours,[2] nor can he be satisfied with those startling Penfield-type cases of experience-brain process correlations performed in surgery.

---

[1] David Hume, who denied that 'we are ever sensible of any connexion betwixt causes and effects' in general and insisted that we have 'only our experience of their constant conjunction' believed that he had thus undercut the traditional Cartesian mind-body problem altogether. But, in our terms, he rejected out of hand only the *IT*, which he considered but another instance of a common epistemological absurdity, and accepted a weak version of *CT* without question. (*Treatise on Human Nature*, Book 1, Section v, p. 247.)

[2] As was done by Hans Berger (1873–1941) in 1928.

(Penfold (1958)). He must be able to specify—with at least reasonable preciseness—the nature of *specific* brain processes and their correlations with experience. Brain *maps*, as they have developed since the turn of the eighteenth century in phrenology and early researches by Franz Joseph Gall (see especially his [1810]) and others (*e.g.* Brain [1964]) are the model for such discovery. Yet no one would dare suggest that we have an adequate 'map' at the present. Thus we are a long way, at least, from a confirmable version of the stronger *CT*.

2 Regarding the stronger version of the *CT*, the proponents of *IT* wish to argue that there are no philosophical considerations which would force us to admit to the impossibility of a future demonstration that every experience is identical to a brain process. But this is not true on at least two counts. First, the move from a strong *CT* to *IT* is not an empirical, therefore not a scientifically demonstrable inference. At best, future scientific investigation will give us a more complete brain map and thus a more defensible stronger *CT*. It will always be a *philosophical*, not a scientific task to move us from the *CT* to *IT*. Even if it is true that *IT* may be more compatible with certain meta-principles of science (simplicity and unity), the argument is not a matter of further investigation and evidence.

Second, the argument for the *IT* falters right at the starting point, for the stronger version of the *CT* can already be argued against, not simply on philosophical grounds, but on scientific grounds as well. There is a notable curiosity involved in the recent debate concerning the *IT*. It is generally agreed that the specifics of the *CT* and the *IT* are not of philosophical concern. In other words, there is no philosophical interest in the *kinds* of correlations discovered by the neurologist in his support of the stronger *CT*, and consequently there has been little interest even in drawing out the distinction between the strong *CT* and the weaker version. Consequently, the neurology discussed by most philosophers is a full three hundred years out of date. Descartes at least knew, what Aristotle did not, that the matter of the brain was the seat of consciousness. This straightforwardly empirical hypothesis is accepted without reservation as central to the *IT* dispute, as is the early discovery that certain specifiable parts of the brain were centres of different sorts of consciousness. It is strange, therefore, that philosophers deny the importance of or remain ignorant of discoveries of recent decades. What is at stake philosophically, of course, is not the specifics of *CT* and *IT* at all, but rather a number of grandiose *Weltanschauungen* which take science as their definitive ideal and outlook. Accordingly, it is a general meta-physics, and not a psycho-neurological theory which holds these philosophers' interest. Still, it is more than curious that this scientific *weltanschauung* takes little interest in the specific advances of the science it defends.

At the turn of the nineteenth century, it was widely believed that mental events could be localised in the brain and detected by formations in the skull. The pop-science of 'phrenology' soon followed these researches, but came under heavy fire from its very beginnings. Hegel took a chapter to make fun of it in his *Phenomenology* ([1807]). Freud precociously argued against it in his early manuscripts and worried about it continuously throughout his life. Even Franz Joseph Gall (1758–1828), generally credited as the founder of phrenology (who is not responsible for the pop-science consequences of his researches), warned against the 'mistake' of theoretical atomisation of brain functions. Phrenology is a simple-minded *CT*. It presumes a one-one correlation of atomic brain events with atomic mental events. Of course, phrenology as a pop-pseudo-science persisted through the nineteenth century, but what is often not recognised is the fact that phrenological models persisted in the science of neurology as well, in fact, up until present day philosophical arguments. *CT* is, I want to argue, a new version of phrenology, and inadequate for reasons argued by Gall and Hegel over a century and a half ago.

At the turn of the present century, it was widely believed that the brain registered and 'stored' 'ideas' (experiences, memories, *etc.*) in specific places. The nerve cell or neurone, discovered in 1838 by Purkyne and so-named by W. Waldeyer in 1891, provided a convenient 'place' for such activity. It was estimated that the average human brain contains in the order of 10 billion neurones. Accordingly, the idea of separate cells, each with its own function and mental *contents*, gained an appeal which persists in much ordinary lay thinking about brain activity to the present. Every experience would be correlated with the stimulation and discharge of a specific nerve cell. (It is not uncommon to find children who report their belief that they can only retain 10 billion ideas and experiences in the same awesome tone that adolescents often believe that they have a given number of sexual climaxes.) Accordingly, *CT* in its strong version awaits only an accurate mapping of the function and contents (or type of possible contents) of each nerve cell, much as a shopkeeper might take inventory of his shelves and storeroom.

This 'atomic theory' of brain-experience localisation has always been a tempting and dominant model for psychoneurology. It has been complicated far beyond its simple-minded phrenological origins by considerable research, the indisputable upshot of which is that even the simplest brain functions are organic activities that involve far more than the stimulation of single cells. Most dramatically, it has been amply demonstrated that stimulation of different cells or areas of the brain can evoke similar experiences, and that destruction of specific parts of the brain is often followed by an apparent transfer of its functions to other parts, sometimes those other than those normally involved in these functions. Even the famous Penfield experiments in which very specific localised stimulations characteristically evoked particular memories or experiences, however they may appear to support simple atomism, do not distinguish between localised stimulation which evokes a localised process from that which touches off a complex process which may involve the entire brain. This does not mean that there can be no brain maps and no localisations and no strong *CT*, but it does mean that the nature of such maps and localisations and *CT* ought to be viewed with some philosophical suspicion. And as more neurological research continues to complicate our picture of the nature of the simplest brain processes, philosophers ought to be respectfully agnostic in their claims, not only of the particulars of brain functioning but in their examples of its forms (*e.g.* the glib talk about 'a *C*-process in the cerebral hemispheres' and the like; for example, to unfairly but respectfully single out one philosophical author, Cornman [1968]).

Since Sherrington's ([1906]) epoch-making work on nervous pathways, it has often been suggested that it is 'brain processes'—sequences of neurones—that are to be correlated with experience rather than individual neurones. According to Sherrington, all brain processes are to be construed in terms of the stimulation of nervous pathways on the model of the reflex arc, in various and intricate complexes. But even Sherrington saw the dangers of what he also called 'phrenology' in the attempted correlation of such processes and specific brain functions:

> Facts rebut the over-simplified conceptions such as ascribed to separate small pieces of the roof-brain, wedged together like a jig-saw puzzle, separate items of highly integrated behaviour. A special place for arithmetical calculation, a special place for musical appreciation, and so on. Such savour of old 'phrenology'. . . . To state the organisation of the mind in terms of roof-brain activities is a desideratum not in sight.

To be sure, there are nervous activities which typically accompany certain experiences. For example, the stimulation of certain pathways through the acoustic cranial nerve might easily be correlated with my hearing a piano middle *C*. But this is surely not the correlation the *CT* has in mind. In persons with certain nervous disorders, the experience of hearing middle *C* takes place without the stimulation of the acoustic nerve. (One might hesitate to say 'hear' here.) And in persons with certain brain lesions, stimulation of the acoustic nerve need not be accompanied by any auditory experience at all. And so we are pushed back to the mystery of the 'higher centres of the brain', where the suggestion that what goes on there is 'brain processes' in terms of stimulation of specific nervous pathways is no better founded than the older supposition that specific cells are the locus of brain activity. Both views are relics of phrenology and both are made doubtful by recent demonstrations that brain functions are variable and transferable.

Of course, phrenology, like most entrenched theories, dies hard. New suggestions of a chemical foundation of brain activity—with the complex 'intelligent' protein *RNA* taking the central role here as it does in embryology—opens up a new possibility of defence. Many neurologists now consider the *RNA* theory the most plausible theory of memory, and some now hope to correlate certain experiences with certain types of *RNA* 'messages'. But this wild possibility has yet to receive confirmation, and its implications are but a matter of wild speculation. If it were to succeed, atomism and phrenology might be given new plausibility.

The repeated and significant findings of the adaptive ability of brain functions has opened an important *empirical* fault in the *CT* (and consequently in the current versions of the *IT* which are based upon the *CT*). Recent neurology has waged a war against just those models upon which philosophers have placed their faith in 'future neurological researches'. *CT* and several isomorphic theories of recent philosophical speculation have been rejected by John Beloff ([1965]) as new versions of the 'pictorial fallacy', by Smythies ([1965]) as 'the representative Theory of perception'. Wilder Penfield (in his [1954] and [1958]) has shown that even some primitive emotional and sexual reactions have proven to be beyond evocation by local stimulation of brain centres. Work in perception by John Platt and others has demonstrated that there cannot be appropriate receptors and nervous networks to give us anything like a correlation between such a simple activity as perceiving a straight line and any specific brain processes.

Given the state of the science as it stands, all evidence points away from the identification of any specific brain processes—whether construed as stimulation of specific cells or specific pathways—which might be correlated with specific experiences. In view of these discouraging findings, the virtue we find in the best neurologists is extreme humility. Lord Brain, for example, says ([1965])

> As regards the physical events occurring in the nervous system, we have at present only the most elementary knowledge of them at comparatively simple levels of organisation. . . .

More cynically, the biologist Ralph Gerard has said ([1959])

It remains sadly true that most of our present understanding would remain as valid and useful if, for all we knew, the cranium were stuffed with cotton wadding.

Accordingly, one finds it astounding that philosophers gratuitously and glibly refer to 'p-processes' and 'excitation of C-fibres' although they have not a clue how these expressions might be correctly instantiated—as if neurologists no doubt do or will know how to fill them out. It is presumptuous to suppose that any such correlations of experiences and processes or excitation of specific nerve cells or pathways will be found, and it is made increasingly less plausible by the researches that are going on. Thus I wish to argue that the strong version of the *CT* is a scientific realist's daydream. (This is emphatically *not* an argument against scientific realism.) At the present time, arguments for the 'ultimate redundancy' of 'mental events' (Quine [1960]) and the identity of causal roles of brain and mental processes (*e.g.* Lewis [1966]) have no basis in neurological researches. The strong version of the *CT* is attractive only so long as we can identify—or have good reason to suppose that we shall be able to identify—specific 'brain processes'. Such is not now the case.

3 If the stronger version of *CT* appears to rest on an inadequate 'phrenological' model, such is not the case for the weaker version of *CT*. And, for the purposes of those philosophers who are interested in defending materialism or the unity of science, perhaps the weak version of *CT* is sufficient. The weak version need make no commitment whatever to the correlation of experience with specific 'brain processes'. It rests content with the far less specific theory that every experience is correlated with activity of the brain, that is, 'some brain process or other' whether it involve a particular set of neurones and nerve paths or perhaps the entire brain. This thesis may not have the heuristic goading effect of the stronger thesis, but it should serve the purposes of interested philosophers.

The key to the materialist and scientific unitarian positions is that an experience cannot be 'something more than' or 'something over and above' the activity of the brain and central nervous system. It is to provide this key that *IT* has been defended, and it is for that purpose that *CT*, at least in its weakened version, has served as a premise. *IT* has been an attractive theory since Spinoza's *Ethics*, in which he rejected Descartes' two-substance view and replaced it with a two-aspects view, one substance viewed from two perspectives and described in two different languages. The modern view has dispensed with the meta-physical notion of 'substance' and has suggested, as Spinoza could not, that there may yet be a way of 'getting behind' the very different 'aspects' and describing the processes themselves in 'topic neutral language', language which is neither solely neurological nor solely psychological. There are some Australian philosophers (notably Smart and Armstrong) who suggest that this is already a real possibility. I believe that they are correct, but it also seems to me that their defence is weakened by their dependence on *CT*, in its weaker as well as stronger versions. In other words, materialism and scientific unitarianism, do not need *CT* in any of its versions.

The mind-body problem has changed since its seventeenth-century formulations. It is no longer involved with 'mind and body' but now with 'sensations and brain processes'. Descartes was aware that the brain was the 'seat of consciousness', and so he anticipated the modern narrowing of the problem. As the philosophers of the nineteenth-century became increasingly sensitive to certain problems of 'intentionality' they further restricted the claim to 'experiential qualities' rather than experience as a whole. It became obvious, for example, that there could be no physiological correlates as such of the objects of experience and thought, such as 'falsely believing that there is no prime number between 3 and 11' or 'hallucinating an irridescent dog in the corner'. Accordingly, the *CT* and consequently *IT* were reformulated in terms of 'sensations' and 'raw feels'.

But the mind-body problem is subject to regression, or has gone in a full circle, if you like, back to the metaphysical framework of Descartes and Spinoza. As neurologists have increasingly identified 'mental' functions outside the brain, it has become increasingly implausible to speak of isolated brain processes. Philosophically Anthony Kenny in his [1971] has convincingly argued, and most biologists have agreed on quasi-empirical grounds, that it is a mistake to ascribe properties of the organism to parts of the organism. Tom Nagel [1965], in his important piece on 'Physicalism', suggests that there is no reason to talk simply about the brain—we can talk about the body. Moreover, it is clearly never the isolated brain that is at stake here (certain 1950s movies aside) but the brain functioning as an *organ* in an *organism*. Accordingly, *CT* is weakened even further to correlate a person's having an experience and something happening in his body.

In accepting this move, I am clearly ignoring the advances of neurology, not only since Descartes, but since Aristotle, who thought the seat of the soul to be the heart and the brain to be a primitive cooling system. Indeed, no philosopher who entertains *CT* can afford to be ignorant of the progress of neurology, but *CT* is not the way to approach materialism and scientific unitarianism. These are not empirical theses at all, but rather like Kant's 'regulative rules'. Accordingly, these principles do not change with progress in neurology or any other science, and their defence does not turn upon *CT* or any other empirical theses. With regard to materialism, I do not see that contemporary philosophers are in any privileged position, and that the philosophers of the seventeenth century, in their neurological ignorance, were perhaps spared some of the errors to which contemporary philosophers are vulnerable. Descartes and Spinoza were not tempted by the promise of neurological maps, and so they did not formulate arguments which depended upon the scientific validity of 'phrenology' or atomism. They were not particularly awed by the complexity of the brain, and so their arguments did not depend upon (however much they speculated about the brain on the side) any thesis which isolated the brain from the body or from 'extension' in general. Both Descartes and Spinoza were deeply concerned about the relationship between mind and body, in which the brain clearly played a central role. But neither was tempted by *CT*, since they realised that the defence of the broad philosophical principles at stake did not depend in the least upon empirical correlations between thought and the cerebrum, thought and the brain, thought and the kidney, or thought and toe movements.

Our objection to the strong *CT* rested upon a charge of mythologising science, believing that it had confirmed or was in the process of confirming theses which are still a matter of confusion and in some cases have been, if not disconfirmed, rendered highly suspect. The weak version of the *CT* is saved from this mythology by its humility, particularly if it is weakened from its brain process-version to its non-committal body-version. But we are already beginning to see that the weak *CT* is losing steam, and it is slowly becoming the almost empty thesis that an experience is correlated with *something* of other. But the weak *CT* suffers from mythologising on the 'experience' side. Again, the *CT* mythology spawns an ontology of questionable individuals, in this case sensations and 'raw feels'. And without an uncontroversial conception of 'experience', of course, the weak *CT* collapses utterly.

In our initial formulation of *CT*, I expressed in passing a certain discomfort with the notion of 'experience'. The complications which arise with the recognition of the intentionality of experience is one good reason for this discomfort, for it then becomes severely problematic how the 'object' of an experience can be included or excluded from the correlate required for *CT*. For this reason, Sellars has led most contemporary philosophers in the demand that the correlate at stake not be experience as such, but sensations or 'raw feels' ('sensa').

But the notions of 'sensation' and its relatives ('raw feels' 'sense data' 'sensa' 'qualia') have received sharp rebuke in recent years, from Wittgenstein and his many followers, and also from phenomenologists, pragmatists and many philosophers of science. The more docile critiques might attack only certain claims made about sensations, for example their alleged 'incorrigibility' and 'givenness'. But the more hostile and increasingly prevalent arguments attack the very idea of 'sensations'. Consequently, philosophers now seem to be returning to a Cartesian notion of 'experience' which includes the *cogitatio* as well as the *cogito* (with minimum emphasis on 'sensations' and the like) or else they have rejected the notion of 'experience' and have turned to more behaviourist models of mind. The key to the argument, whether in Merleau-Ponty or Sellars, is that, in so far as it is justifiable to talk about 'sensations', they are not *data* at all and not particulars. When Quine glibly argues that mental events are nothing but theoretical entities, and when James Cornman hesitantly replies that they are not theoretical but data of sorts, they well illustrate a point that most philosophers would accept, that the status of sensations can no longer be considered the uncontroversial starting point around which any ontology must be developed. Most philosophers today have moved away from sensations as the *data* of the philosophy of mind and rather concern themselves with the question whether talk of sensations is *necessary*, *i.e.* theoretically defensible. Similarly, most philosophers have left behind the old empiricist idea of sensations as peculiar atoms of experience and have turned to the Wittgensteinian notion of 'having a sensation', or the Sellarsian idea of sensations as 'states of sentient beings, indeed primarily of persons', and neither particulars nor givens. Accordingly, the correlate of *CT* can no longer be simply 'there is an experience', but rather it will be a controversial theoretical claim that this organism or person has a certain sensation. But even here, it can be easily demonstrated that an organism or a person 'having a certain sensation' is not to be mistaken for an analysis of everyday experience. 'Having a sensation', a headache, a pain in a gouty toe, or a sensation of dizziness, are intrusions into our everyday experience, not its fundamental units. Consider your sitting on the grass, sipping a glass of wine while listening to a friend read a poem. It would be manifestly absurd to suggest that such a complex, yet comparatively simple experience could be broken into discrete units of 'sensations'. It is important to stress the fact that any such description of everyday experience refers not only to some 'mental condition' of the subject but to his position in the world and the 'objects' of his experience, the grass, the wine, the poem and its connotations. (*Cf.* Nagel [1965] sect. IV.) However one might attempt to 'sensationalise' such experiences, it is clear that the result will always be less than the total experience itself.[1] Thus the apparently uncontroversial suggestion that there are discrete units on the 'mental' side of the mind-body problem is itself a highly dubious—if not outlandish doctrine.

---

[1] We cannot help but notice the similarity between this claim and that of several leading '*Gestalt*' psychologists. Curiously, however, the *Gestaltist* school used this same claim in the defence of a highly speculative and since proven implausible thesis of mind-body 'isomorphism', a relatively strong version of *CT*. (See, for example, Köhler [1930].)

4 It is often claimed that *CT* is free of ontological commitment, that it is an empirical claim which is open to any number of ontological interpretations. Thus it is supposed that *CT* in its weak version is a well-confirmed empirical hypothesis, while *CT* in its stronger form is a partially confirmed outline of a system of hypotheses, the details of which will be filled in with further observation. It is uncommitted, it claims, to the number of entities or events involved—whether there are two independent kinds of entities or events ('substances'), two causally related entities or events, or a single set of entities or events viewed from different aspects or described in different languages. But *CT* is neither a simple empirical claim nor is it ontology-free. It abstains from only one ultimate ontological decision—between monism and dualism, but it is brutal in its exploitation of the ontological groundwork within which the mind-body/sensations-brain process problem is formulated. The *CT* carves up its world into sensations and brain processes. Both of these are controversial ontological commitments, not simply the data from which an ontological problem arises. Consequently, the *CT* is neither a pure empirical thesis nor is it free of serious ontologising.

The problem with *CT*, in all its versions, is its intrinsic dualism, even when it is to serve as the basis for a monistic *IT*. It begins by carving up the world into mental and physical, and then attempts to carve up each of those realms into particulars that can be matched with particulars on the other side. But, if the point of both materialism and the unity of science is ontological monism, beginning with a dualistic *CT* is a most awkward way of proceeding.

University of Texas, Austin
and
University of Michigan, Ann Arbor

REFERENCES

Baier, K. [1962]: 'Smart on Sensations', *Australian Journal of Philosophy*, **40**, pp. 57–61.

Beloff, J. [1965]: 'Comments' in J. R. Smythies (*ed.*): *Brain and Mind*, pp. 192–5.

Brain, W. R. [1965]: 'Some Aspects of the Brain-Mind Relationship', in J. R. Smythies (*ed.*): *Brain and Mind*, pp. 63–78.

Brandt, R. [1960]: 'Doubts About the Identity Thesis', in S. Hook (*ed.*): *Dimensions of Mind.*

Cornman, J. [1968]: 'On the Elimination of "Sensations" and Sensations', *Review of Metaphysics*, **22**, p. 16.

Gall, F. J. [1810]: *Anatomie et Physiologie.*

Hegel, G. W. F. [1807]: *Phänomenologie des Geistes*, in Schulze (*ed.*): *Werke*, pp. 237–54.

HUME, D. [1732]: *Treatise on Human Nature*, Selby-Bigge (*ed.*).

JASPER, H., PENFIELD, W. [1954]: *Epilepsy and the Functional Anatomy of the Human Brain.*

KENNY, A. [1971]: 'The Homoculus Fallacy', in M. Grene (*ed.*): *Interpretations of Life and Mind.*

KÖHLER, W. [1930]: *Gestalt Psychology.*

LEWIS, D. [1966]: 'An argument for the Identity Theory', *Journal of Philosophy*, **43**, pp. 17–25.

NAGEL, T. [1965]: 'Physicalism', *Philosophical Review*, **64**, pp. 339–56.

NIETZCHE, F. [1882]: *Beyond Good and Evil.*

PENFIELD, W. [1958]: 'The Role of the Temporal Cortex in Recall of Past Experiences', in Wolstenholm and O'Connor (*eds.*): *The Neurological Basis of Behaviour.*

PLATT, J. [1963]: 'How We See Straight Lines', *Scientific American*, 1963.

QUINE, W. V. [1960]: *Word and Object*, pp. 264–6.

SHERRINGTON, C. [1906]: *The Integrative Action of the Nervous System.*

SMYTHIES, J. R. [1965]: 'The Representative Theory of Perception' in J. R. Smythies (*ed.*): *Brain and Mind*, pp. 241–56.

Brit. J. Phil. Sci. 26 (1975), 41–60 *Printed in Great Britain*

Discussions

CONSCIOUSNESS, MIND, AND SELF: THE IMPLICATIONS OF THE SPLIT-BRAIN STUDIES

Introduction.

1 *An Alternative Account to Puccetti's.*

2 *What of Persons?*

Introduction

In his recent essay Puccetti ([1973]) has provided a thorough and interesting treatment of the many fascinating issues now emerging from the phenomenon of the bisected human brain. His analysis was, in my view, quite successful in marshalling the arguments against the standard interpretations of these experimental findings. However, I think his final resolution is unsatisfactory as well. In the present essay I would like to offer an alternative account (to all the previous interpretations) which I believe is more successful in reconciling the divergent views at issue.

The central points at issue have been: How many minds may the split-brain subject be said to possess; and as a corollary, how many persons are to be associated with a given individual?[1] Prior to Puccetti's arguments the disputants arranged themselves into three distinct camps:

> (1) Those who maintain that these individuals have two distinct minds (one associated with each hemisphere) and are nevertheless only one person (*e.g.*, Sperry, Bogen, Gazzaniga, *et. al.*). According to this position a person is something distinct from his mind; thus we can have two minds but only one person.
>
> (2) Those who hold that these individuals have exactly one mind (associated with the major hemisphere). (Eccles). Presumably Eccles would answer the question of the number of persons involved by likewise holding that only one person may be predicated of each individual. And since for Eccles only the major hemisphere has a mind we would, presumably, associate the person with his left brain. Or in any case Eccles seems committed to denying that there are *two* persons here (unless of course he countenances mindless persons).
>
> (3) Those who hold that while these individuals certainly have only one mind, it is inappropriate to predicate a mind of either hemisphere (MacKay here). MacKay, along with Eccles, would thus seem committed to rejecting the notion that more than one person is involved (unless of course *he* countenances mindless persons).

[1] I shall use 'individual' as a neutral term to indicate one human body with its associated minds and persons (however many there may be).

Puccetti's position is in agreement with Sperry *et. al.* in holding that these individuals have two minds; but since he apparently does not countenance *personless minds*, he concludes that these individuals really are two persons as well.

1 *An Alternative Account to Puccetti's*

The thesis I will advance here is a synthesis of some aspects of the previous interpretations. I intend to side with Eccles and MacKay in insisting that only one person is involved here; and to side with Puccetti and Sperry in admitting what seems to me the undeniable evidence for two minds at work in these subjects. In particular I will try to forge a synthesis between Eccles' interpretation and that of Sperry *et al.* In this way I hope to do fuller justice than Eccles to the compelling evidence regarding the cognitive powers of the minor hemisphere; *and* to incorporate Eccles' insight regarding the *decisively important* difference language makes for the assessment of the relative status of the two hemispheres. My contention is that Eccles has erred in refusing to acknowledge how impressive the minor hemisphere's cognitive powers really are; and that the Sperry group (and Puccetti) have erred in failing to recognise that the possession of language makes for a *profound* disparity between the two cerebral hemispheres. In particular I maintain that the presence of language marks the difference between the presence of *self-consciousness* and the complete absence of any awareness of self.

In order to appreciate my contention we must first draw more sharply the distinctions between the notions of 'consciousness', 'mind', and 'self'. As I see it, these distinctions unfold in a hierarchical manner. At the lowest level we have consciousness; by consciousness I mean the basic phenomenal awareness that accompanies acts of perception, emotions, sensations, *etc.* I hold that mind is something over and above this. In order to possess a mind the organism must in addition possess a hierarchical ordering of behavioural priorities which are consciously utilised in intentional actions. That is, in order to have a mind an organism must be able to formulate intentional behavioural plans, priorities of action.

As examples of organisms at the first level (conscious without a mind) I take all lower organisms whose behaviours are primarily reflexive, as in various tropisms and purely instinctive reactions. I would say that an earthworm, for example, is conscious of various sensations as a result of heat, moisture, *etc.*, but that its reactions to stimulation are purely reflexive and tropistic—the earthworm does not form intentions to move away from light, it does so purely on the basis of a negative phototropism. As examples of organisms at the second level (conscious and mindful but devoid of self) I take all those higher organisms whose behaviour apparently exhibits intentionality. For example a cat does not climb a tree as a function of any tree-climbing tropism, but rather it seemingly formulates a certain intention regarding various behavioural options open to it.

But even above both these levels is the level at which we find *consciousness of self.* By consciousness of self I mean the ability to apprehend oneself as a being distinct from other similar beings; to recognise one's actions and thoughts as 'belonging' in some sense to oneself. That one's thoughts are one's own, as they say, is an expression of this awareness. We might speculate that with the emergence of self-consciousness in the evolutionary development of the hominids we can mark the transition from ape to man. This transition might be betokened by the appearance of burial customs (around the time of Neanderthal). Certainly it is the awareness of oneself as a unique being, distinct from one's fellows, and limited in space and time, which makes for the awareness of death. Self-consciousness, then, would seem to play a major role in the evolution of man. As Dobzhansky has stated it:

> Self-awareness is, then, one of the fundamental, possibly the most fundamental, characteristic of the human species. This characteristic is an evolutionary novelty; the biological species from which mankind has descended had only rudiments of self-awareness, or perhaps lacked it altogether. The self-awareness has, however, brought in its train sombre companions—fear, anxiety, and death-awareness (Dobzhansky [1967], p. 68).

Now my contention is that the minor hemisphere is deficient in just this respect; that it lacks the very awareness of self which is so integral in making the human mind what it is. And my conjecture is that it is the presence of language which allows self-awareness to appear. Certainly it is language which, more than anything else, separates man from his fellow primates—it is language which allows man to achieve the levels of cognising which infrahuman animals apparently lack. And certainly it is language which allows man to attain the level of abstraction necessary to apprehend the pathos of death.[1] However I think this level of self-awareness is vastly elaborated, and that self-awareness appears at a more rudimentary level. Recent evidence indicates, for example, that the higher apes may possess a more rudimentary form of self-awareness (this will be discussed below). For this crude form of self-awareness we require not necessarily evidence of death-awareness, but only evidence that the animal can apprehend itself as a distinct being, separate from its fellows. That it grasps the notion of its own personal separation from its environment and its compeers.

The studies bearing on self-awareness in the higher apes were undertaken by Gallup (cf. Gallup [1970], [1971]). Gallup studied the ability of chimpanzees to differentially respond to mirror images of themselves vs. mirror images of other chimps. Initially a chimpanzee responds to a mirror in a social fashion, *i.e.*, as if another chimp were present. Gallup found that after a few days exposure to the mirrors the chimps learned to associate the image with their bodies and begin to use the mirror to groom themselves, manipulate various body parts, *etc.* It was also found that monkeys fail to exhibit this type of learning—after prolonged exposure they still respond to their mirror images as if confronting another animal.[2]

Seemingly what has occurred is that the chimps begin by responding to their mirror images as being those of other animals; then gradually they learn to associate the image with themselves in a manner which seems to betoken apprehension of the 'self' concept. Now the point to notice about these studies is that they demonstrate the possibility of testing for self-awareness in a manner not requiring verbal communication. And a further important point is that monkeys are unable to respond to their mirror images in this way. This fact coupled with the recent success at teaching non-verbal languages to chimpanzees (and the inability of monkeys to learn these languages) would lend corroborating support to the conjecture that language is essential for self-awareness.

---

[1] Karl Popper has recently argued for just this same view of the relationship of language to consciousness and self-consciousness (cf. Schilpp [1974], pp. 150–2). Popper develops his view on the basis of his metaphysical theories regarding the 'three worlds'. He also explicitly mentions the split-brain studies in this context; although not as an interpretation of those studies but rather as a possible test arena for his theory.

[2] Gallup reports no results with any of the other great apes.

Now I certainly believe that the minor hemisphere lacks anything like full human self-awareness; I suspect, for example, that it does not agonise over its inevitable death the way its companion hemisphere does. In this sense, then, I am inclined to say that it is less than human. Whatever the case may be, we can never ascertain whether the minor hemisphere is deficient in this manner, owing to its muteness. However, I am willing to conjecture that not only does the minor hemisphere fail to achieve full human self-consciousness, but that it even fails to achieve the level of self-awareness exhibited by the higher apes! Since this level of self-awareness is apparently testable through purely non-verbal behavioural evidence, the mute hemisphere may have a fair chance of evincing whatever level of self-consciousness it may have.

My interpretation of the split-brain studies, then, goes as follows: Both minor and major hemispheres are conscious in that they both, no doubt, have the basic phenomenal awareness of perceptions, sensations, *etc.* And they both have minds (under my definitions) in that they exhibit elaborated, organised systems of response hierarchies, *i.e.*, intentional behaviour. But in addition I would conjecture that only the major hemisphere has a self; only the language utilising brain is capable of the abstract cognising necessary in order to be aware of itself as a unique being. In a word, only the major hemisphere is aware of itself *as a self*.[1]

Hopefully this conjecture may be amenable to possible experimental test, provided we can specify objective behavioural criteria for attribution of selfhood. Now I do not think that the specification of such criteria is as hopeless a task as it may at first appear. Harkening back to Gallup's work with the self-concept in chimps, it seems possible to construct an experimental design in such a way as to provide intuitively plausible behavioural criteria which give evidence of the presence of self-awareness. Admittedly the problem is a bit more recondite when we are inquiring after the self-awareness of a mute half-brain. In any event what I would like to urge is that a concerted attempt be made to ascertain if the present hypothesis can possibly be tested. As far as I can determine from the literature, such an attempt has never been made.

However, even if the present interpretation should prove refractory to experimental check I would still urge its superiority as a strictly philosophical account of the split-brain phenomena. It seems to me a clear advance over Eccles' view in that Eccles has repeatedly denied that the minor hemisphere is even conscious. But the evidence indicates quite clearly that the minor hemisphere is both conscious and possessing of mind (in my terms). On the other hand Puccetti, in company with Sperry *et. al.*, would have us believe that the two hemispheres are on roughly equal ground; it is just that they each exhibit certain specialised talents absent from their partner. Thus the major hemisphere is said to be analytic, verbal, philosophical (!), while the minor hemisphere is intuitive, spatial, musical, artistic, *etc.*[1]

---

[1] Please note that I do not mean to implicate the existence of *more* than two spheres of awareness. The self subsumes what I have called the 'mind' and the 'consciousness' in the major hemisphere; the mind likewise subsumes what I have called 'consciousness' in the minor hemisphere. In each case the relations unfold in a hierarchically inclusive fashion. In all, then, there are only two spheres of awareness, one of which is richer than the other in that it is *self-aware*.

Eccles, then, would have us believe that the difference between the two hemispheres is just that between consciousness and unconsciousness. Puccetti would have us believe that the differences are just those between partners of equal status with differing talents—rather like being proficient at golf *vs.* being proficient at tennis. Both views, I suggest, are too extreme. We must, it seems to me, grant the impressive cognitive abilities of the minor hemisphere (it simply won't do to say it is unconscious). On the other hand we need to do justice to the difference that possession of language makes (this difference, *pace* Puccetti, must be seen as an important one). I believe that the present account satisfies both these desiderata and is thus preferable to any previous interpretations of this data.

2 *What of Persons?*

As I stated at the outset two questions have been at issue here: (1) How many minds may the split-brain individual (and the intact-brain individual as well) be said to possess?; and (2) How many persons are associated with each individual? My answer to the first query is that these individuals possess precisely *two minds*, one associated with each hemisphere; but that the major hemisphere is much richer in that it also possesses a self. Now my answer to the second question is, 'It depends on what you mean by "person"'.

Apparently Puccetti takes what I am calling 'mind' as equivalent to person. However, I am inclined to argue that the absence of a self in the minor hemisphere (if I am correct) is sufficient to disqualify it as a person. It seems to me that we are wont to attribute personhood to individuals who exhibit something more closely approximating to self-awareness. I would be inclined, for example, to deny personhood to one's cat on the ground that it is not a self-aware being. I would grant, of course, that it is conscious and possesses a mind but that it simply fails to possess a self; and thus, I would say, is not a person. I would urge that a person is not to be identified with a certain body's conscious experiences but only with a certain subset of those experiences, the subset containing those conscious experiences which are experiences of *self*.

---

[1] Ornstein carries this type of line to its most extreme. He maintains (cf. Ornstein [1972]) that the two hemispheres constitute two fundamental and distinct modes of knowing. The one is analytical, verbal, linear and rational (the major hemisphere), and the other is intuitive, non-verbal, non-linear, a-rational (the minor hemisphere). He then suggests that this distinction corresponds, respectively, to the Western and the Eastern philosophical approaches to reality.

I should point out that Ornstein's rather grand interpretation is not necessarily incompatible with the present construal. Indeed the most salient characteristic of most Eastern-type mystical or meditational states of consciousness is precisely *absence of sense of self.* If, as Ornstein suggests, we are to understand the difference between the Western and Eastern 'modes of knowing' as that between the functioning of the major and minor hemispheres, then absence of sense of self is precisely what the present account would predict for such Eastern-type experiences. However Ornstein's rather bold conjecture will have to bear closer scrutiny before we can assess its cogency.

The answer I would offer to the query regarding the number of persons involved is the same answer I would give to the question of how many selves an individual possesses—*one*, associated with the major hemisphere. The left brain is thus the person I refer to by the name 'John Doe'; there is no more nor no less of John than there ever was. John's right brain now becomes a part of John much as John's arm or leg can be said to be a part of John (albeit in this case a *conscious* part). Previously our naive tendency was to associate John with his whole brain; the splitting of John's brain makes this association untenable. We erred before not so much in associating a person with a brain, but only in associating him with *too much* of a brain!

Puccetti invites us to consider the problems occasioned by any view predicating single personhood of these individuals. We seem, for example, to be forced to say that in those cases wherein one hemisphere knows something that the other does not, that the person both knows and does not know something at the same time. Now of course this seems contradictory. But it is only formally a contradiction if we identify the person with the knower, *i.e.*, if we can say: a knower (*K*) both knows and does not know some proposition *P* at time *T*. And of course ordinary language makes this identification of the person with the knower seem quite natural. The force of my view would be to deny that the person can be identified with the knower, *tout à fait*. Rather we must recognise that not only can a person know things but a *part* of the person (his minor hemisphere) can also know things. Thus whenever we refer to a given act of knowing it is an open question as to whether or not the person (*i.e.*, the major hemisphere) knows the information in question.

Admittedly this does radical violence to our ordinary language discourse regarding knowing. But then the entire bisected brain phenomenon does radical violence to a great many of our ordinary conceptions of things! I am inclined to opt for a reformation of our ordinary discourse regarding persons, selves, minds, knowing, *etc*. However, the problem is partly semantical and I really care very little whether we are all of us really two 'persons'. I do however intend to insist that the two hemispheres are on radically unequal ground; that the major hemisphere is richer in an important sense than its minor relative. And if we decide that we are all of us really two persons then I shall merely say, 'Fine! But let us just remember that one of us is a much more important person than the other!'.

LARRY W. DEWITT
Indiana University

(For the sake of bibliographic completeness I have included a few significant items not contained in Puccetti's original references.)

REFERENCES

Dobzhansky, T. [1967]: *The Biology of Ultimate Concern*.
Eccles, J. C. [1967]: 'Evolution and the Conscious Self', in J. D. Roslansky (*ed.*): *The Human Mind*, Third Nobel Conference.
Eccles, J. C. [1968]: 'The Experiencing Self', in J. D. Roslansky (*ed.*): *The Uniqueness of Man*, Fourth Nobel Conference.

Eccles, J. C. [1973]: 'Brain, Speech and Consciousness', *Naturwissenschaften*, **60**, pp. 167–76.

Gazzaniga, M. S. [1972]: 'One Brain—Two Minds?', *American Scientist*, **10**, pp. 311–18.

Gallup, G. G. Jr. [1970]: 'Chimpanzees Self-Recognition', *Science*, **167**, pp. 86–87.

Gallup, G. G. Jr. [1971]: 'It's Done with Mirrors—Chimps and Self-Concept', *Psychology Today*, 4 (10), pp. 58–61.

Kimura, D. [1973]: 'The Asymmetry of the Human Brain', *Scientific American*, **10**, pp. 70–78.

Nagel, T. [1971]: 'Brain Bisection and the Unity of Consciousness', *Synthese*, **22**, pp. 396–413.

Ornstein, R. [1972]: *The Psychology of Consciousness*.

Puccetti, R. [1973]: 'Brain Bisection and Personal Identity', *British Journal for the Philosophy of Science*, **24**, pp. 339–55.

Schilpp, P. A. (*ed.*) [1974]: *The Philosophy of Karl Popper*.

THE INTRANSITIVITY OF NON STANDARD SYNCHRONISMS

Introduction.

1 *Quinn's Argument.*

2 *The Flaw in Quinn's Argument.*

Introduction.

In a recent article Quinn[1] has questioned the validity of the proof offered by Grunbaum[2] that all linear rules $t=t_1+\epsilon(t_2-t_1)$, $0<\epsilon<1$, for the setting of distant clocks are intransitive if $\epsilon\neq\frac{1}{2}$ and offered a counter example which he claims is both transitive and need not have $\epsilon=\frac{1}{2}$. I will show that this counter example is, in fact, also intransitive. The flaw in Quinn's argument is that he has not, in fact, tested his procedure for transitivity, only applying it to one degenerate case.

Before proceeding it is essential to have a clear understanding of transitivity of synchronism. I take the definition of Reichenbach[3]:

If two clocks C_1 and C_2 are set from a clock A by some rule r, then if C_1 and C_2 are directly compared with each other by the same rule r then they will be found to agree.

Let us denote the setting of clock C_1 to agree with clock A by rule r as C_1S_rA. The operation S_r is then transitive if C_1S_rA and C_2S_rA implies $C_1S_rC_2$.

It is equally important to have a clear definition of the rule r for setting clocks; in Grunbaum's (or Reichenbach's) case the rule is clearly defined; clock B is set equal to clock A if a light-signal leaving A at time t_1 on A's clock reaches B at a time t_2 on B's clock and returns to A at a time t_3 on A's clock, where $t_2=t_1+\epsilon(t_3-t_1)$.

1 *Quinn's Argument*

Consider three clocks A, B, C at rest in an inertial system, let the distances AB, BC and AC, be l, m and n respectively. The clock B is set relative to clock A by the rule r: namely $t=t_1+\epsilon(t_2-t_1)$. The clock C is then set by the same rule

[1] Quinn [1974].

[2] Grunbaum [1969].

[3] Reichenbach [1927], p. 168.

r relative to B. Quinn then demonstrates that the clock C is set relative to A by the same rule r, *provided the light signal traverses the path ABC.*

The proof follows by considering a light ray leaving A at t_0 on A's clock, reaching B at t_1 on B's clock, the signal is partially reflected back to A reaching A at time t_2 on A's clock, and partially redirected to C arriving at time t_3 on C's clock, where it is reflected back to B arriving at t_4 on B's clock and finally redirected back to A at time t_5 on A's clock. Since BS_rA and CS_rB we have

$$t_1=t_0+\epsilon(t_2-t_0),\ t_2-t_0=2l/c \tag{1}$$

$$t_3=t_1+\epsilon(t_4-t_1),\ t_4-t_1=2m/c \tag{2}$$

$$t_5-t_4=t_2-t_1=(1-\epsilon)2l/c. \tag{3}$$

Elementary algebraic manipulation yields

$$t_3=t_0+\epsilon(t_5-t_0),\quad t_5-t_0=2(l+m)/c. \tag{4}$$

Hence Quinn concludes that CS_rA and hence the rule r is transitive.

2 *The Flaw in Quinn's Argument*

As I emphasised in the introduction it is necessary to have a rule r for setting clocks. In Quinn's case the light ray to set a clock C_1 must go via B. Our S_r sign means that any other clock is set by a light signal that first goes to clock B thus we should amend our S_r sign to S_{r_B}. Our rule r_B is now transitive if

$$C_1S_{r_B}A \text{ and } C_2S_{r_B}A \text{ implies } C_1S_{r_B}C_2.$$

We thus need to consider another clock C_2.

Let the distance AB, BC_1 and BC_2 be l, p and q respectively. Let $C_1S_{r_B}A$ and $C_2S_{r_B}A$ with arbitrary ϵ. Consider now a light signal leaving C_2 at time t_1 on C_2's clock, reaching B at time t_2 on B's clock, deflected to C_1 arriving at time t_3 on C_1's clock, returning to B at time t_4 on B's clock, and deflected to C_2 arriving at time t_5 on C_2's clock. Then since

$$C_1S_{r_B}A \quad t_3=t_2+2p/c \qquad t_4=t_3+(1-\epsilon)2p/c \tag{5}$$

$$C_2S_{r_B}A \quad t_2=t_1+(1-\epsilon)2q/c \qquad t_5=t_4+\epsilon 2q/c \tag{6}$$

a little manipulation yields

$$t_3=t_1+2\epsilon p/c+(1-\epsilon)2q/c. \tag{7}$$

On the other hand if $C_2S_{r_B}C_1$ then

$$t_3=t_1+(t_5-t_1)=t_1+\epsilon(2p+2q)/c. \tag{8}$$

These equations are clearly only compatible if either $\epsilon=\frac{1}{2}$, or if $q=0$. Quinn's analysis corresponds to $q=0$ so that clock C_2 is identical to clock B. In the general case C_2 is any other clock and the rule r_B is transitive only if $\epsilon=\frac{1}{2}$.

Indeed, it is obvious that this must be the case for if the rule S_r is transitive so that C_1S_rA and C_2S_rA implies $C_1S_rC_2$, replacing C_1 by A and C_2 by B gives AS_rA and BS_rA implies AS_rB so that S_r is symmetric. The only linear rule $t=t_1+\epsilon(t_2-t_1)$ that is symmetric has $\epsilon=\frac{1}{2}$.

I. W. ROXBURGH
Queen Mary College
University of London

REFERENCES

GRÜNBAUM, A. [1969]: 'Simultaneity by Slow Clock Transport in the Special Theory of Relativity', *Philosophy of Science*, **36**, pp. 5–43.

QUINN, P. L. [1974]: 'The Transitivity of Non Standard Synchronisms', *The British Journal for the Philosophy of Science*, **25**, pp. 78–82.

REICHENBACH, H. [1927]: *Philosophie der Raum-Zeit-Lehre*. Dover Translation *The Philosophy of Space and Time*, 1956.

ON THE MACH PRINCIPLE AND GENERAL RELATIVITY

In a recent note in this *Journal*,[1] Woodward and Yourgrau discussed some ideas that get to the heart of current controversies on the subject of the Mach principle *vis à vis* the theory of general relativity. I should like to add a few remarks in this note on points where I especially feel there to be misunderstanding in their paper.

1 The first misunderstanding I wish to discuss is Woodward's and Yourgrau's claim, based on my [1972*a*] that it is my 'view that the bulk of the inertia of an object is created locally by vacuum polarisation effects'—'a view that is almost undoubtedly false in view of the results obtained through mass and charge renormalisation in quantum electrodynamics'. As to the first part of this quotation, what I was referring to in my [1972*a*] as the major local contribution to the inertia of an elementary particle, is the matter of the *real* electron-positron pairs (needed in any case to explain many of the other observed physical properties of the microscopic domain of matter). This influence of the background pairs for any observed elementary particle manifests itself in terms of the curvature of space-time—which their existence implies, in accordance with theory of general relativity. With this theory, the curvature of space-time was shown to relate directly, in an explicit mathematical fashion, to the inertia of an observed elementary particle. Thus, the mass of the elementary particle is an explicit function of the other matter of the assumed closed system of general relativity theory—in full accord with the requirement of the Mach principle. A point that I emphasised in my [1972*a*] was that the contribution of the distant stars to the curvature field at the site of the observed electron is of negligible importance in comparison with the contribution to this field from the pairs of the background. Thus the Mach principle is being exploited in this derivation, but without the need to introduce astronomical factors (a view that Mach himself believed to be the case!). The details of the theoretical development[2] were referred to in my [1972*a*]. Had the authors checked this reference, they would have seen that my result has nothing to do with quantum electrodynamics or the renormalisation method—which in fact are rejected, both conceptually and mathematically, in the field theory I refer to. Nor is my result related in any way to the other works they cited, by Tomozawa, and by Sakurai, which do relate to quantum electrodynamics.

I might add that in a later paper,[3] the mass of the electron (and the muon, which appears as a member in a mass doublet) was calculated numerically, in accordance with the formal structure of this theory, and both masses were found to be in close agreement with the experimental values. Thus, in contrast with the remarks of Woodward and Yourgrau, a full exploitation of the theory of general relativity, ***which incorporates the Mach principle in an explicit mathematical fashion***, that is indeed mathematically consistent, does lead to an observed fact of nature. In accordance with the commonly accepted philosophy of science, if these authors wish to reject my result in a scientific way, they must technically refute the mathematical derivations that I have published in my [1968] and [1972*b*].

---

[1] Woodward and Yourgrau [1973].

[2] Sachs [1968].

[3] Sachs [1972*b*].

2 Even if I had been referring to the physical model and the computational techniques of quantum electrodynamics, as Woodward and Yourgrau falsely assumed I was doing, the latter part of their statement, quoted above, is clearly a *non-sequitur*. For the physical model of vacuum polarisation has never been shown to be unambiguously implied in a demonstrably consistent form, either logically or mathematically, by the computational techniques of renormalisation. Thus, one can hardly draw conclusive statements from quantum electrodynamics that would necessarily rule out the results of a different theory—especially if the latter is logically and mathematically consistent!

3 The authors assert that because 'the Schwarzschild metric does not asymptotically approach the Minkowski metric everywhere as the radius of the single field source is contracted to zero . . . we may thus conclude that Mach's principle and general relativity are incompatible'. This conclusion is fallacious. The reason is that the (exterior) Schwarzschild solution that the authors refer to is not actually singular in the regions of space where it is a valid solution. Recall that this solution solves a restricted form of Einstein's field equations (with an imposed spatial spherical symmetry) in which the matter source field, $T_{\alpha\beta}$, is approximated by zero. But the singularity in this solution occurs where $T_{\alpha\beta}$ cannot be approximated by zero (inside of the matter that is giving rise to the geometrical field). Unlike the case of linear field theories, such as Maxwell's equations for electromagnetism, the solutions of the field equations, $R_{\alpha\beta}=0$, that Schwarzschild took to replace Einstein's field equations, $R_{\alpha\beta}-\frac{1}{2}g_{\alpha\beta}R=T_{\alpha\beta}$, are not generally contained in the family of solutions of Einstein's equations. This is due to the nonlinear structure of these partial differential equations.

Neither is it true, as the authors contend, that according to general relativity theory the space-time must necessarily be flat at infinity. This boundary condition was originally imposed in the Schwarzschild problem in order to match the (empirically confirmed) lack of influence of a single centre of gravitational force, such as the sun, on a test body that is infinitely far away from it. But such a boundary condition is not built into the theory of general relativity! It is only used in special cases as a useful approximation. Since, according to general relativity theory, any amount of matter in the universe implies that space-time is curved where that matter is, the tensor property of the curvature field then implies that space-time is curved *everywhere*. To replace a small section of curved space-time, at an infinitely distant point from some assumed centre of force, by the tangent (flat) space-time at that point, is to use a useful approximation. But it is important to recognise that the tangent space-time (in which one may use the Minkowski metric of special relativity) is not, in principle, contained in the actual curved space-time.

4 Finally, I should like to express full agreement with the authors' comment that 'Mach's principle is still of fundamental importance in gravitational theory'. However, I do not agree with their rejection of the unified field approach, and therefore believe that it is logically necessary to extend the Mach principle to all of the physical manifestations of matter. I am convinced of this not only by the argumentation based on the logical content of the full conceptual basis of the theory of general relativity, but also by the concrete mathematical results that have been achieved in my research in incorporating this theory with the Mach principle, as demonstrated in my [1968]. In this regard, recall Einstein's advice[1]: 'If you want to find out anything from the theoretical physicists about the methods they use, I advise you to stick closely to one principle: don't listen to their words, fix your attention on their deeds.'

MENDEL SACHS
*State University of New York
at Buffalo*

REFERENCES

EINSTEIN, A. [1954]: *Ideas and Opinions.*

SACHS, M. [1968]: 'On Positive Definiteness of Inertial Mass, Nonpositive Definiteness of Electromagnetic Coupling and The Mach Principle from General Relativity', *Nuovo Cimento*, **53B**, pp. 398–414.

SACHS, M. [1972*a*]: 'On the Mach Principle and Relative Space-Time', *British Journal for the Philosophy of Science*, **23**, pp. 117–21.

SACHS, M. [1972*b*]: 'On the Electron-Muon Mass Doublet from General Relativity', *Nuovo Cimento*, **7B**, pp. 247–64.

WOODWARD, J. F. and YOURGRAU, W. [1973]: 'Mach's Principle, The Equivalence Principle and Gravitation: A Rejoinder to Newburgh', *British Journal for the Philosophy of Science*, **24**, pp. 264–9.

[1] Einstein [1954], p. 270.

MYTHS AND X-RAYS*

1 *Introduction.*
2 *The Standard History.*
3 *Forman's Criticisms of the Standard History.*
4 *What Ever Happened to Space Lattices?*

1 *Introduction.*

I shall appraise two rival accounts of von Laue's discovery of X-ray diffraction, with particular reference to the role of the 'space lattice' theory of crystal structure in this discovery. The two accounts are:

(1) The standard history of the discovery; the salient points of which were

---

* I would like to thank Grant Barnes, Gregory Currie, Peter Grant, K. Long and especially Peter Clark for the help that they gave me in preparing this discussion note. Above all I would like to acknowledge the great intellectual debt that I owe to my teacher, the late Imre Lakatos.

already contained in von Laue's Nobel lecture,[1] which I shall take together with similar accounts by Bragg[2] and Ewald[3] as my sources.

(2) Forman's account[4]; which can be considered as an attempt to supersede the traditional 'internal' history by an 'external' one,[5] in this case one which relies upon psychological conjectures about Ewald's character.

I shall show that while the traditional history is certainly inadequate, the problems which Forman's account attempts to solve are the products of a naive methodology, and vanish when the correct reconstruction of the contemporary problem situation is given. In this sense, the case history illustrates the primacy of 'internal' over 'external' history.

2 *The Standard History.*

The standard account of the discovery of X-ray diffraction proceeds as follows:

From their discovery in 1895 by Röntgen, the nature of X-rays became a matter of major controversy. In particular Röntgen's discovery gave rise to much research into the relationship between X-rays and the various known forms of radiation. Together with electromagnetic radiation, in particular light in the visible spectrum, the mysterious 'Röntgen rays' exhibited polarisation, scattering by matter (already observed by Röntgen himself), and the photoelectric effect. However until von Laue's discovery of 1912 all attempts to display diffraction had proved at best inconclusive, though such phenomena were to be expected if X-rays were electromagnetic in nature.

However if X-rays were not a manifestation of electromagnetism, as the absence of diffraction phenomena seemed to indicate, what were they? If X-rays *could* be diffracted, then it would show that the interpretation of X-rays as electromagnetic radiation was a plausible one,[6] since it would mean that X-rays exhibited, at least from a qualitative point of view, the same phenomena as electromagnetic radiation, whatever the nature of *that* radiation might turn out to be.

The search for X-ray diffraction goes back to Röntgen himself. Later experiments, involving, like Röntgen's, the use of a slit, were performed by Haga and Wind and Walter and Pohl. Some success was claimed in both cases, though in the former the experimental result was challenged on the grounds that the observed phenomena might well have occurred during the development of the photographic plate used in the experiment.

---

[1] Von Laue [1920].

[2] Bragg [1933].

[3] Ewald [1962].

[4] Forman [1969].

[5] I am using the terms 'internal' and 'external' in the sense of Lakatos [1972]. Lakatos takes internal history as the intellectual history of the growth of scientific theories—written in the light of a specific methodology. External history refers to the 'empirical (socio-psychological)' history with which any internal history 'needs to be supplemented' (Lakatos [1972], p. 91). I wish to support here the claim of Lakatos, that internal history is the more fundamental of the two categories, in so far as sociology and psychology is necessary to explain only those parts of the history of science which are not appraised as rational by the specific methodology adopted.

[6] Bragg and von Laue himself are much more extreme. Von Laue writes that the diffraction of X-rays 'solve [d] the question of whether or not [X-rays] represent a wave phenomena or the ejection of . . . small particles' (von Laue [1920], p. 347), whilst Bragg writes, 'It was not until 1912, when von Laue showed [an X-ray] could be diffracted like ordinary light, that it was recognised with certainty as [a] . . . wave' (Bragg [1933], p. 228).

However the experiment of Walter and Pohl led, through measurements by Koch and their interpretation by Sommerfeld, to the following claim: If X-rays *were* waves, and if the Walter–Pohl photographs *were* caused by diffraction phenomena, then X-rays must have a wavelength of about 4×10^{-9} c.m. (*i.e.* approx. $10^{-3} \times$ wavelength of visible light). But if X-rays were really of such small wavelength, it would be technically impossible to construct any known form of diffraction grating adequate to demonstrate this diffraction unambiguously. Such a construction would have involved the drawing of millions of parallel lines to an inch!

In 1912, von Laue was working in Sommerfeld's institute for theoretical physics in Munich. Also present at the institute was Ewald, who was researching into the behaviour of long electromagnetic waves passing through a crystal. During a conversation with Ewald about the nature of Ewald's research it occurred to von Laue that space lattices (the regular three dimensional arrays of atoms, of which, according to the Munich physicists, crystals were supposed to consist[1]) might well provide an ideal diffraction grating for X-rays. Von Laue realised that the spacings of the lattice would be the right size to cause the diffraction of X-rays, if indeed such a phenomenon existed.

Acting on von Laue's idea an experiment was designed and performed by two of his Munich colleagues, the experimental physicists Friedrich and Knipping. This experiment[2] was 'completely successful'[3] in demonstrating the diffraction of X-rays by a crystal. This provided a strong confirmation of the view that X-rays were merely electromagnetic waves, akin to visible light, but of a considerably shorter wavelength.

This account is straightforward enough. It is, roughly speaking, an 'internal' account in as much as it refers only to theories and experiments and the logical relations between them. But often such an account needs to be supplemented by 'external' history, as interesting questions of a social or psychological nature may need to be answered. Forman has posed such a question, and he claims that any acceptable answer to it shows up the standard history as inadequate. First I shall present this problem and outline what the standard history has to say about it. Then I shall present Forman's criticisms of the standard history and his own answer to the problem.

All this was in 1912, while the space lattice theory had reached its high-water mark twenty years earlier with the independent compilation by Schoenflies and Fedorow of a list of all (230) possible space groups.[4] Why then had von Laue's idea not occurred to someone before? Why, with such a complete theory of crystal structure, did it take so long for anyone to perform an experiment of the kind performed by Friedrich and Knipping? Ewald tells us that: 'In spite of the elaborately worked out theory, the physical significance of these geometrical constructions [*i.e.* the space lattices of Schoenflies and Fedorow] remained obscure'.[1] But what then was it about the intellectual atmosphere at the University of Munich in 1912 that made their physical significance cease to be obscure? Or to pose the question in the manner of Forman, 'Why Munich?'[2]

---

[1] According to the space lattice theory, all crystals are built up by the translational repetition of 'motif structures' (groups of atoms, in modern versions of the theory, exhibiting a particular symmetry and generated by the repetition of 'motif units' in accordance with this symmetry) on 'space lattices' (three dimensional networks generated by unit cells—for example, a cubic lattice derives from a cubic unit cell). For a clear mathematical account of how the symmetry of the motif structure is connected with the symmetry of the unit cell and other related matters, *cf.* Jaswon [1965].

[2] Reported in Friedrich, Knipping and Laue [1912].

[3] Bragg [1933], p. 235.

[4] The totality of operations which generate all the motif units of a crystal from any one of them constitute the 'space group' of the crystal. It can be shown that such space groups are in fact groups in the mathematical sense, *i.e.* they satisfy the axioms of group theory. It can also be shown (*cf.* Jaswon [1965], especially p. 39) that the number of different symmetries and hence the number of space groups, which can be exhibited by any crystal is finite.

Let me say at once that the question that Forman raises, 'why Munich?' is an essentially *external* question, and I shall not therefore try to answer it in purely internalist terms. But such socio-psychological problems are best dealt with when treated *in the light of* a correct methodology. In fact my answer to Forman will consist primarily of an attempt to correct his own, perhaps tacit, methodological presuppositions.

In his Nobel lecture von Laue cites confidence in the space lattice theory at Munich at the time when it subsisted in other research centres 'in the form of a questionable hypothesis ... a somewhat unknown quantity to physicists'[3] as being one of the major theoretical influences that led him to his idea. In this claim he has the retrospective support of Ewald[4] and several modern textbooks.[5] It was this confidence, together with the 'active advocacy of the wave theory of X-rays'[6] which supposedly made the intellectual atmosphere at Munich uniquely favourable to the devising of the Laue–Friedrich–Knipping experiment.

3 *Forman's Criticism of the Standard History.*

According to Forman, von Laue knew of the space lattice theory *before* his conversation with Ewald, indeed *before* he ever came to Munich.[7] The space lattice theory, Forman argues, was well known outside of Munich, and his evidence is most persuasive.

(1) Only eight of the twenty-three textbooks on mineralogy and crystallography published between 1890 and the time of the experiment, fail to mention the space lattice theory. A further two books, while giving mention to the theory,

---

[1] Ewald [1962], p. 29.

[2] Forman [1969], p. 41.

[3] Von Laue [1920], p. 351. Von Laue speaks from a more or less justificationist standpoint, and is not agreeing with Popper, according to whom *all* scientific theories are questionable hypotheses.

[4] *Cf.* Ewald [1962], p. 24, where he states that the space lattice theory had been 'discredited', and pp. 41–2 where he claims the honour of having been the first to inform von Laue that at Munich, 'crystals were thought to have internal regularity'.

[5] *Cf.* for example, Lipson ([1970], p. 17), who writes that before his consultations with Ewald (and consequently prior to his arrival in Munich) 'Laue knew nothing of the [space lattice] ideas about crystals'. Azároff ([1968], p. 76) also stresses the importance of von Laue's initiation into the mysteries of the space lattice theory by Ewald. Von Laue himself is less extreme. In emphasising the importance of the Munich environment, he acknowledges his debts for learning the theory to the Munich crystallographer P. Groth, as well as to Ewald (von Laue [1920], p. 351). However the consensus seems to be that von Laue had no knowledge of the theory of space lattices prior to his coming to Munich.

[6] Forman [1969], p. 41. I do not intend to discuss here this other claimed intellectual influence at Munich, which is discussed and criticised by Forman ([1969], pp. 52–64).

[7] This would not be too surprising anyway, for von Laue had studied under Voight at Göttingen.

take a phenomenological stance and do not make any theory of crystal structure the basis of their development of crystallography.[1]

(2) Although crystallography was not an integrated part of physics at that time, most German physics and mathematics students studied it to some extent, for reasons connected with the structure of the physics curriculum and the German educational system. They will then have become acquainted to some extent with the theory of space lattices through the books referred to in (1).[2]

But now two problems arise:

(A) If the space lattice theory was more pervasive than suggested by the standard history, how could von Laue write, only eight years after the discovery, that the space lattice theory prior to the discovery of X-ray diffraction, 'remained a somewhat unknown quantity to physicists'?[3]
(B) How did this 'myth' come to be perpetuated right up to the present day?

Forman proposes socio-psychological solutions to both problems:

According to him, von Laue invented the story of the inadequacy of the space lattice theory in order to detract from his own part in the discovery. He was led by his modesty to give the credit for his idea to the 'environment' rather than to his own genius. For Forman then this part of the standard history constitutes the 'wholly fictional rationalization of a man who regarded himself as lacking in originality'.[4]

As for the second question—the peculiar tenacity of the myth—Forman argues that what he considers the essentially falacious traditional account of the discovery, functions as a sociological device to 'maintain a separate identity'[5] for the community of X-ray crystallographers. Separate, that is, from other fields that use its techniques: 'physics, chemistry, crystallography, geology and now biology'.[6] Myths in general and this myth in particular fulfil their function of 'strengthen-[ing] tradition' by 'tracing [the tradition] back to a higher, better, more supernatural reality of initial events'.[7] In this case the aspect of the traditional history which is saddled with the tags 'higher, better' and 'supernatural' is the suggestion that this and other scientific enterprises may involve something like an objective rationality.

Forman seems to want the history of science to concern itself primarily with the sociology and psychology of the scientific community. For instance, Forman states that, 'it is perhaps possible to . . . consider the process of mythicization itself in the hope of finding guidelines for interpreting those statements by scientists which purport to be historical'.[8] And in a footnote to this, Forman writes that this 'problem is similar to that facing the historian of primitive societies'.[9] Is science then to be regarded as just another belief system, epistemologically on a par with the beliefs of primitive societies? Perhaps science as the traditional paradigm of rationality should not be sacrificed with so much alacrity. At least in this case, I shall show that there is no need to do so.

---

[1] For a full breakdown of this statistic, *cf.* Forman [1969], footnote 40, p. 47.

[2] *Cf.* Forman [1969], pp. 48–9.

[3] Von Laue [1920], p. 351.

[4] Forman [1969], p. 42.

[5] Forman [1969], p. 40. Forman is not clear whether the attempt to 'maintain a separate identity' was a conspiracy by X-ray crystallographers or merely an unconscious tendency.

[6] Forman [1969], p. 40. [7] Forman [1969], p. 67. [8] Forman [1969], p. 68.

[9] Forman [1969], footnote 116, p. 68. The comparison of primitive societies and scientific communities is presumably more than just an analogy.

4 *Whatever Happened to Space Lattices?*

I want to show in this final section that von Laue's account of the role of the space lattice theory is not sufficiently inaccurate to warrant Forman's externalist story.

The theory that the 'distinctive property of the crystalline state is its ultimate internal periodicity',[1] dates back to Haüy who held that crystals were built up by the repetition of ultimate crystalline units, which he called *molécules intégrantes*. As to the exact nature of these basic units he was not at all sure. However early work in the theory of elasticity by Navier, Poisson and most especially Cauchy, led to the view that the theory was 'a brilliant, but unfortunately not acceptable speculation which neither physicists nor crystallographers dared to use seriously'.[2] I shall now describe these early developments.

By a generalisation of Hooke's Law due to Cauchy[3] we obtain:

$$F = Cx$$

where F is a 6-vector whose elements are the sheering and normal stresses on a crystal lattice, x is a 6–vector of corresponding strains and C is a 6×6 matrix of elastic constants.

The *particular* space lattice theory of Poisson and Cauchy predicted that there would be 15 independent elastic constants in the most general case, though only one for isotropic bodies. This yields Poisson's ratio as 1/2 for a stretched rod. This so called 'rariconstant' programme was challenged by the 'multiconstant programme' of Green, Stokes and William Thompson, which was not based on the underlying structure of crystalline matter. They predicted that 21 of the elastic constants would, in general, be independent, with 2 independent in the isotropic case. This meant that Poisson's ratio could be any number at all.[4] Measurements by Wertheim in 1848 'raised serious doubts that Poisson's ratio was always equal to 1/2'[5] while the 'non-validity of Cauchy's relations [was] convincingly demonstrated by Woldemar Voight's measurements of elastic deformations of anisotropic crystals'.[6]

Despite this chaos in the camp of Cauchy, we must agree with Forman that the theory of space lattices seems to have survived this 'refutation' surprisingly well. But as Forman takes pains to point out, the space lattice theory was never an issue in the argument between the rariconstantists and the multiconstantists. Indeed, not only was it not challenged by the multiconstantists, but it was actually accepted by them though they made no use of it for calculating Poisson's ratio, or for making other deductions in the theory of elasticity.[7]

I want to maintain that all this in no way prevents us from agreeing with Ewald in his assertion that, 'it seemed probable that the Cauchy relations did not hold, and this opinion discredited the model from which the Cauchy relations sprang, namely that, in the natural state of a crystal its molecules are arrayed in a three-dimensional lattice'.[1] Forman, however, does not accept this. He writes:

---

[1] Ewald [1962], p. 22.

[2] Ewald [1962], p. 24.

[3] *Cf.* Ewald [1962], p. 23.

[4] *Cf.* Forman [1969], pp. 44–6, for details of this controversy. *Cf.* also Todhunter and Pearson [1886], I, pp. 496–506.

[5] Forman [1969], p. 45.

[6] Ewald [1962], p. 24.

[7] *Cf.* Forman [1969], pp. 45–6.

> 'This myth [*i.e.* of the special status of the space lattice theory at Munich] . . . attained its fullest elaboration, replete with quite fictitious details, in Ewald's retrospective account on the occasion of the 50th anniversary of the discovery of X-ray diffraction. Now, rather than simply being without physical consequences and unknown, the space lattice hypothesis is said to have been experimentally refuted. The relations between the elastic constants which Cauchy had deduced *ca.* 1830, for a lattice of identical point centers of force "were not confirmed by experiment, and the failure discredited the entire concept of internal regularity of crystals".'[2]

In fact all that Ewald has claimed is that the problems surrounding the Cauchy relations lead to the 'discrediting' of the space lattice theory. Indeed the word 'discredited' appears both in my quote from Ewald[3] and in a similar quote that Forman extracts from another of Ewald's publications. Thus, implicit in the long quote that I have taken from Forman is the equation:

$$\text{Discredited} = \text{Rejected}$$

On this view, one anomaly constitutes a refutation—a reason to reject the theory—even when no one has put forward a better theory. But science has rarely operated like this. Even the best scientific theories develop in an 'ocean of anomalies'.[4] In this sense, all scientific theories are 'discredited'. However they may remain unrejected for centuries. The question is rather, does a theory make novel predictions which are subsequently corroborated?

Bearing this methodological point in mind, let us take a final look at the situation of the space-lattice theory prior to the Laue–Friedrich–Knipping experiment.

The space-lattice theory, it is true, was not rejected. But what sort of existence did it have after the work of Wertheim and Voight?

On the basis of a *particular*[5] space-lattice theory, the rariconstantists had made an interesting novel prediction about the value of Poisson's ratio, but this prediction turned out to be an anomaly for the theory. The multiconstantist Kelvin showed that this anomaly could be removed by a simple, though *ad hoc*, modification of Cauchy's assumptions. This led to the independence of 21 elastic constants in the general case and 2 in the isotropic case, in accordance with the multiconstantist programme. Of course this modification meant that one could no longer make any prediction as to the value of Poisson's ratio, for no such prediction could be made on the basis of two independent elastic constants for isotropic bodies. Even what I have referred to as the 'high-water mark' of the space-lattice theory[6]—the calculation of the complete list of space groups by

---

[1] Ewald [1962], p. 24. Ewald, of course, just uses the term 'probable' in an ordinary language sense, and not in the sense of the probability calculus.

[2] Forman [1969], p. 43.

[3] *Cf. above*, p. 54, note 4.

[4] *Cf.* Lakatos [1970], pp. 133, 138. The methodological ideas presented here are derived from the writings of Imre Lakatos.

[5] The fact that Cauchy's own version of the space lattice theory turned out to be unsatisfactory does not mean that the idea of space lattices was an incorrect one. All that may be indicated is the need to replace some of the minor assumptions in Cauchy's deductions. This is what Born later did.

[6] *Cf above* p. 53.

Schoenflies and Fedorow—was very much an exercise in pure mathematics. Ewald claims that: 'In spite of the elaborately worked out theory, the physical significance of these geometrical constructions remained obscure'.[1] In particular 'What was the nature of the particle inserted in, and multiplied by, the framework of symmetry elements'?[2] In his Nobel lecture, von Laue had earlier painted a similar picture[3] and Ewald has defended this position against Forman's strictures, stressing 'the vagueness of ... Bravais' and Schoenflies–Fedorov's [space lattice theories, which] is shown by the fact that nobody could tell what physical entity should be placed in the equivalent points.[4] Indeed no 'far reaching physical conclusion had evolved from this line of thought'.[5] It sufficed only to explain the most elementary of crystalline properties, for example the law of rational indices. Nevertheless the space lattice theory was accepted even by those phenomenologically inclined scientists like Voight, whose work helped to put the nail in the coffin of the rariconstant programme; accepted that is because no alternative had been put forward. According to the methodological theory that I have adopted here,[6] it would clearly have been irrational to reject space lattices until a better theory came along. The multiconstantist programme which did not base its predictions on any theory of the internal structure of crystals was in no position, of course, to provide any alternative to the space lattice theory. The multiconstantists were merely concerned with the elastic properties of crystals. However the space lattice theory was, to a great extent, a dormant programme. Little research was being carried out into its intricacies. It was only under the impetus of the Laue–Friedrich–Knipping experiment and discovery that the programme got going again. In particular Max Born re-examined Cauchy's deductions and found that they contained a flaw.

Only at Munich was there actual *advocacy* of the space lattice theory, as opposed to its mere mention in textbooks. Even though the space lattice hypothesis had few physical consequences, in Munich Sohncke's crystal lattice models were freely available and the crystallographer P. Groth actually spread the gospel of space lattices. Laue himself writes, '... in Munich where models of the Sohncke space lattices were to be found in more that one university institute, it was P. Groth who expressed his defence of [the space lattice theory] both orally and in writing, and I ... learned [it] from him'.[7] And over thirty years later he still tells the same story, '... the university had a collection of models from the times of Leonhardt Sohncke, who had taught in Munich until 1897 and had made considerable contributions to the mathematical theory of space lattices ... Paul von Groth used the crystal structure models in his lectures'.[8]

There are two points to be made here. Firstly, methodologically speaking, there is nothing wrong with working on a research programme which is in as poor shape as the space lattice theory seemed to be, just so long as its inadequacies are admitted by its advocates.[9]

Secondly, and more important in this context, though both the von Laue quotes above are from defences of the thesis that Munich had a peculiarly stimulating intellectual atmosphere for work in this field—a thesis which Forman contests—Forman does not reject these actual assertions about the situation at Munich.

---

[1] Ewald [1962], p. 29.
[2] Ewald [1962], p. 29.
[3] Von Laue [1920], p. 351.
[4] Ewald [1969] p. 74.
[5] Von Laue [1920], p. 351.
[6] *Cf. above*, p. 52, note 5.
[7] Von Laue [1920], p. 351.
[8] Von Laue [1951], p. 293.
[9] *Cf.* Lakatos [1972], note on page 104.

He just claims that they give no significant support for the case of uniqueness.[1]

What may we conclude from all this? There seem to be two rival historical hypotheses under discussion:

(1) The space lattice theory—which once was a theory of great promise—was refuted and rejected by the scientific community, when Wertheim and more particularly Voight demonstrated its inconsistency with the facts. However, at Munich the community of physicists, crystallographers and mineralogists, just ignored the refutation[2] and went on talking about space lattice as if nothing had happened. Von Laue, knowing little or nothing about the space lattice theory, came to Munich and became acquainted with the theory after talking to Groth and Ewald and after seeing Sohncke's models. This led him to devise the famous diffraction experiment—later performed by Friedrich and Knipping—which through its success, led to the correction of Cauchy's space lattice theory and to its resurrection.

(2) The space lattice theory—which was a theory of great promise—had a severe set back when Wertheim and more particularly Voight showed that deductions, on the basis of Cauchy's particular version of the theory, which seemed to lead to novel predictions, turned out to be anomalous. Faced with no alternative theory, the scientific community payed lip service to the theory of space lattices. The one exception was Munich, where the great lattice theorist Sohncke had worked and where Groth kept his memory alive and was encouraging research in this field. Given that Munich was the place where actual research on space lattices had been done, one would naturally expect scientists there to take an active interest in the fate of the theory, even though its progress had been halted.

Thus the question 'Why Munich' can be answered at least in part, by substituting for 'Munich', 'the place where scientists working on the programme were gathered'. Now, as I pointed out above, this does not entirely rule out externalist considerations. We could still ask, why were space lattice researchers working at Munich, rather than somewhere else? Questions like this *always* remain after an internal explanation has been given. Anyway, this is not the problem that we are dealing with here.

Von Laue, before coming to Munich treated the theory in the same contemptuous manner as his non-Munich colleagues. Once at Munich, however, he came under the influence of Groth. In discussions with Ewald he came to realise the potential of the space lattice theory for diffraction esperiments. This led him to devise his famous experiment—later performed by Friedrich and Knipping—which through its success led to the correction of Cauchy's space lattice theory and to this theory's resurrection.

Hypothesis (1) has been argued against—in my opinion quite conclusively—by Forman. However I want to put forward hypothesis (2) as being an adequate internal account[3] of the role of the space lattice theory in the discovery of X-ray diffraction and one which does not deviate too widely from the standard accounts.

Unfortunately, Forman does not seem to have noticed the difference between (1) and (2). Thus he seems to believe that the term 'discredited' can mean nothing other than 'refuted and rejected'. This becomes clear, when, in answer to his own question: 'What then was the conceptual situation in physics *circa* 1911?'[1] he claims: 'Contrary to the Laue–Ewald thesis, but as the foregoing discussion[2] would lead us to anticipate, the existence of the space lattice was taken by the physicists as a matter of course; almost no one even thought to label it as an assumption.'[3]

---

[1] *Cf.* Forman [1969], p. 42.

[2] For some unexplained reason.

[3] *Cf. above*, p. 52, note 5.

In his reply to Forman, Ewald is quick to point out that von Laue, in his Nobel lecture, only claimed a favourable climate for Munich, not a *uniquely* favourable one. That is to say, Ewald is giving his support to hypothesis (2) rather than hypothesis (1). Ewald writes: 'Nowhere [in the Nobel lecture] does Laue represent Munich as '*extraordinarily if not uniquely*' favourable to the conception of his experiment. This aggrandizement is left to [Forman] who needs it in order to present a "myth" which he can destroy'.[4]

We can see then that Forman's problems arise out of the inadequacy of his philosophy of science to deal with the problem situation of the time. He does not realise that a more sophisticated methodology would enable him to appraise the traditional history more favourably, rendering his external socio-psychological explanation superfluous.

L. D. GASMAN
London School of Economics

REFERENCES

AZÁROFF, L. [1968]: *Elements of X-ray Crystallography.*

BRAGG, W. [1933]: *The Universe of Light.* (The page number refers to the Dover reprint, 1959.)

EWALD, P. P. [1962]: *Fifty Years of X-ray Diffraction.*

EWALD, P. P. [1969]: 'The Myth of Myths; Comments on P. Forman's paper on "The Discovery of the Diffraction of X-rays in Crystals",' *Archive for History of Exact Sciences*, **6**, pp. 72–81.

FORMAN, P. [1969]: 'The Discovery of the Diffraction of X-rays by Crystals; A Critique of the Myths', *Archive for History of Exact Sciences*, **6**, pp. 38–71.

FRIEDRICH, W., KNIPPING, P. and LAUE, M. von [1912]: 'Interferenz-Erscheinungen bei Rontgenstrahlen', *Bayerische Akademie der Wissenschaft zu München*, pp. 302–22. (The page number refers to the English translation in G. E. Bacon: *X-ray and Neutron Diffraction*, pp. 89–108, 1968.)

JASWON, M. A. [1965]: *Mathematical Crystallography.*

LAKATOS, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. E. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*, pp. 91–195.

LAKATOS, I. [1972]: 'History of Science and its Rational Reconstructions', in R. C. Buck and R. Cohen (*eds.*): *Boston Studies in the Philosophy of Science*, **8**, pp. 91–135.

LAUE, M. von [1920]: 'Uber die Auffindung der Röntgenstrahlinterferenzen', *Les Prix Nobel en* 1914–1918. (The page numbers refer to the English translation in *Nobel Lectures: Physics 1901–22*, pp. 347–55.)

LAUE, M. von [1951]: *My Development as a Physicist.* (The page numbers refer to the English translation in Ewald [1962], pp. 278–307.)

LIPSON, H. S. [1970]: *Crystals and X-rays.*

TODHUNTER, I. and PEARSON, K. [1886]: *History of the Theory of Elasticity and of the Strength of Materials.* (The page number refers to the Dover reprint, 1960.)

---

[1] Forman [1969], p. 49.

[2] That is Forman's arguments against hypothesis (1).

[3] Forman [1969], p. 49.

[4] Ewald [1969], p. 73.

Brit. J. Phil. Sci. 26 (1975), 61–71 *Printed in Great Britain* 61

Review Article

THE STRUCTURE OF SCIENTIFIC INFERENCE*

1 *Hesse's Problem: a Seeming Paradox of Transitivity.*

2 *Hesse's Solution: Argument by Analogy and Clustering Postulates.*

3 *Apparent Counterexamples to Hesse's Solution.*

4 *A Theorem in Probabilistic Confirmation Theory which Seems to Dissolve Hesse's Problem and Hesse's Solution.*

5 *Hesse's Argument for the Irrelevance of Theories.*

6 *Objections, Logical and Psychological, to Hesse's Argument for the Irrelevance of Theories.*

7 *Hesse's Unwillingness to Assign Finite Probabilities to Universal Generalisations.*

8 *Good Reasons for Assigning Finite Probabilities to some Universal Generalisations.*

9 *The Propriety of Assigning Finite Probabilities to each of an Open-Ended Infinity of Universal Generalisations over Infinite Domains.*

10 *Need we Adopt the Bayesian Prescription for Changing our Beliefs?*

11 *The Applicability of Convergence-to-Opinion Theorems to the Case of Actual Scientific Inference.*

12 *Hesse's Claim to be a Bayesian Personalist.*

Suitable as a text for advanced graduate students, this book brings together as a unity the arguments that Hesse has published in a number of papers in recent years, and aims to make a major contribution to our understanding of induction and in particular of the role of theories in mediating inductive inferences.

1 *Hesse's Problem: A Seeming Paradox of Transitivity.*

Central to Hesse's whole discussion is what she calls 'the transitivity paradox', namely the fact that if the evidence e_1 increases the probability of a hypothesis h, and if h entails a further evidential statement e_2, it is not in general the case that e_1 thereby also increases the probability of e_2. As a counterexample let e_1 and e_2 be any evidential statements whatever and take h as the conjunction of e_1 and e_2. Hesse sees her central problem as something like the following: How is it that, when h is a scientific theory, we nevertheless, apparently justifiably, take evidence for this theory as increasing the probability that other predictions of the theory will be confirmed?

---

* Review of M. B. Hesse [1974]: *The Structure of Scientific Inference*. London: Macmillan. £5·95. Pp. vii+309.

2 *Hesse's Solution: Argument by Analogy and Clustering Postulates.*

Hesse proposes the following solution to this problem: a scientific theory, in order to transmit probability in the way we require, must express an antecedently present, but not necessarily recognised, real *analogy* between the predictive instances in question. In virtue of this analogy we are enabled to infer, from the co-presence of certain properties in the examined instances, their co-presence in the unexamined instances as well. The type of confirmation theory which is required in order to legitimise this type of probabilistic inference is, Hesse argues, one in which the assignment of prior probabilities satisfies a *clustering* postulate which will state, crudely-speaking, that properties that go together in one case tend to go together in every case. Hesse sees arguments by analogy/clustering postulates as lying behind every type and principle of legitimate scientific inference.

3 *Apparent Counterexamples to Hesse's Solution.*

Is there not, however, something wrong with the analysis of the problem situation which this solution presupposes? Consider just about the simplest situation that any theory of scientific inference must cope with, namely one where two *rival* theories, h and g, agree in predicting e_1 but predict e_2 and $-e_2$ respectively. We cannot have transitivity of confirmation in this situation without getting into contradiction. It cannot therefore be the case that *in general* evidence for a theory does increase the probability that other predictions of the theory will be confirmed.

Let us consider, for example, Tycho's and Copernicus's theories, which agree in predicting the phases of Venus but disagree about the possibility of observing stellar parallax. We cannot have transitivity of confirmation both to Tycho's prediction and to Copernicus's prediction because these are incompatible. Is Hesse's analysis supposed, perhaps, to discriminate between the two predictions, to tell us, in cases of conflicting theories, which theory is to be preferred? But in this example I cannot see how either theory can be plausibly construed as expressing an analogy between its respective pair of predictive instances. This would only seem possible when both predictions at issue can be construed as particular instances of some single unifying law embodied within the theory. But that seems far too special a case of a scientific theory 'justifiying inductive inference from the evidence to further predictions' for Hesse to be able to claim to have given a general solution of the problem she posed initially.

But other examples seem to do even more serious damage to Hesse's thesis. Consider the case of Einstein and Ritz agreeing in predicting the null result of the Michelson-Morley experiment but disagreeing about what would be the observed velocity of light emitted from a source moving rapidly relative to the observer. Einstein construes both his predictions as instances of the law that the velocity of light is constant relative to any observer. Ritz construes both his predictions as instances of the law that the velocity of light is constant relative to the light source. I do not see how Hesse could claim to detect a significant analogy in the one case but not in the other. Yet we cannot have transitivity of confirmation in both cases without getting into contradiction.

I do not think these are isolated counterexamples. Take the case of Newton and Huygens agreeing in predicting Snell's law but disagreeing about the relative velocity of light in dense media. The bending of the incident ray towards the normal as it enters the denser medium is in Huygens's theory an instance of the velocity of light decreasing in the denser medium, but in Newton's theory it is an instance of its increasing as it enters the denser medium. If Hesse cannot construe either of these theories as asserting significant analogies between their respective pairs of predictive inferences, but only rather forced formal analogies, then how can her analysis cover the general case of scientific inference? Or take the case of the Steady-State Theory in cosmology construing both the universal Hubble recession and the equality of mean density of galaxies, whatever their age or distances, as two instances of the universe's appearing the same to any observer at any time, while its rival the Big-Bang Theory construes the greater mean density of more distant, and hence older, galaxies as simply a special case of the general expansion of the universe with time, of which the simplest evidential instance is the observed Hubble recession. Here again, if we take relations of analogy implicit in the respective theories as justifying the 'required' transitivity of confirmation, then we simply get a contradiction.

The conclusion we ought to draw from these examples is surely that we do not require the kind of transitivity of confirmation for scientific theories and their predictions that Hesse's programme is designed to produce.

4 *A Theorem in Probabilistic Confirmation Theory which Seems to Dissolve Hesse's Problem and Hesse's Solution.*

Not only are relations of analogy apparently not sufficient to do the job but they seem to me also not necessary for any job that needs doing. All we need seems to me already given by the following theorem of ordinary probabilistic confirmation theory. Let e_1 be the original evidence, e_2 the novel prediction; let h be the theory we are interested in and which predicts e_2 (*i.e.* $p(e_2, h) = 1$); let h' be the disjunction of all other serious hypotheses which agree with h in predicting e_2 ($p(e_2, h') = 1$); let g be the disjunction of all other serious hypotheses which predict $-e_2$ ($p(e_2, g) = 0$). I assume that this partition exhausts all serious relevant hypotheses: that is to say if some serious relevant hypothesis i does not satisfy either $p(e_2, i) = 1$ or $p(e_2, i) = 0$, then I assume that we disjoin it into two more specific hypotheses i_h, and i_g, such that $p(e_2, i_{h'}) = 1$, $p(e_2, i_g) = 0$. Using approximate equality ($\simeq$) to indicate that other contributions to the subjective probability are subjectively negligible in the present context, we then have:

1. $p(h) + p(h') + p(g) = p(h \vee h') + p(g) \simeq 1,$
2. $p(e_1) \simeq p(e_1, h \vee h')p(h \vee h') + p(e_1, g)p(g),$
3. $p(e_2) \simeq p(e_2, h \vee h')p(h \vee h') + p(e_2, g)p(g) = p(h \vee h') = p(h) + p(h'),$
4. $p(e_2 \,.\, e_1) \simeq p(e_2 \,.\, e_1, h \vee h')p(h \vee h') + p(e_2 \,.\, e_1, g)p(g) = p(e_2 \,.\, e_1 \,.\, h \vee h')$
$$= p(e_2, e_1 \,.\, h \vee h')p(e_1 \,.\, h \vee h') = p(e_1 \,.\, h \vee h').$$

From 4 we obtain:

5. $p(e_2, e_1) = p(e_2 \,.\, e_1)/p(e_1) \simeq p(e_1 \,.\, h \vee h')/p(e_1) = p(h \vee h', e_1) = p(h, e_1)$
$+ p(h', e_1)$.

Comparison of 3 and 5 will give us one set of necessary and sufficient conditions for transitivity of confirmation. But it is also convenient to derive an equivalent set expressed in terms of g:

From 1 and 3 we obtain:

6. $p(e_2) \simeq 1 - p(g)$.

Using 5 and then 2 we obtain:

7. $p(e_2, e_1) \simeq p(e_1, h \vee h')\,(p(h \vee h'))/p(e_1) \simeq (p(e_1) - p(e_1, g)p(g))/p(e_1)$
$= 1 - p(e_2, g)/p(e_1) = 1 - p(g, e_1)$.

From 6 and 7, we have

$$p(e_2, e_1) \geqslant p(e_2) \text{ iff } p(g, e_1) \leqslant p(g);$$

from 3 and 5, we have

$$p(e_2, e_1) \geqslant p(e_2) \text{ iff } p(h, e_1) + p(h', e_1) \geqslant p(h) + p(h').$$

In other words we have transitivity of confirmation from e_1 to e_2 via h if and only if e_1 does not significantly reduce the probability of h' (the serious rival hypotheses which agree in predicting e_2) by as much or more than it significantly increases the probability of h. Or equivalently e_1 will significantly increase the probability of e_2 if and only if it significantly reduces the probability of g (the serious hypotheses which make the opposite prediction).

Since the result is exactly what we expect for scientific inference, I hold that there is no 'paradox of transitivity' but merely a *fallacy of transitivity*, which arises whenever, with Hesse, we attempt to ignore the effect of the evidence on serious rival theories.

5 *Hesse's Argument for the Irrelevance of Theories.*

Hesse does give an independent argument for the irrelevance of theories to instance confirmation, an argument which she seems to regard as simply another formulation of her 'paradox of transitivity'. This runs (p. 144):

> Expressed in Bayesian language the paradox is represented as follows. We are interested in the value of $p(e_2, e_1)$, where e_1 is an observed or otherwise given consequence of a theory h, and e_2 is an as yet unobserved or otherwise problematic further consequence of h, that is, an untested prediction made by h. Now $p(e_2, e_1)$ is a single-valued probability function of its arguments alone, and its value cannot depend on whether or not we are interested in some particular h from which e_2 and e_1 are deducible. e_1 and e_2 are in general deducible from a number of different sets of premises in the language, whether we have thought of them or not. Some of them are perhaps potential theories. But whether they turn out to be viable as theories or not, $p(e_2, e_1)$ has its unique value independent of them.

(In her next paragraph, Hesse promises to make this argument 'a little more transparent in probabilistic terms' but then merely disposes (or does she?—I leave her readers to judge) of an ingenious (but incorrect) possible objection to her thesis.)

Hesse later (p. 176) summarises this section:

> The non-transitivity of confirmation has led to the conclusion that if we are concerned with confirmation of predictions from hypotheses . . ., then the probability of hypotheses is irrelevant to confirmation of predictions.

6 *Objections, Logical and Psychological to Hesse's Argument for the Irrelevance of Theories.*

Contrary to the impression given by the passage from Hesse, because of the requirements of coherence—that is to say in virtue of the axioms of the probability calculus, the value we assign to $p(e_2, e_1)$ is ***not independent*** of the values we assign to such other probability functions as $p(e_2, h)$, $p(h, e_1)$, $p(e_2, -h)$, $p(-h, e_1)$. $p(e_2, e_1)$ does not have a 'unique value independent of' the available relevant theories, for the value I assign to it has to be coherent with the degrees of belief that I attach to them.

Consider Hesse's own favoured example. Let e_1 be the (approximate) validity of Kepler's laws and let e_2 be the prediction that there will be a departure (of a certain specified kind) from Galileo's law of falling bodies near the earth's surface. Now suppose that I am impressed by Huygens's view that the orbits of the moon and planets are caused by the axial rotations of their central bodies (which view, unlike Newton's, explains why the orbits lie more or less in a single plane, why the central bodies exhibit rotations in the same direction as their orbiting satellites, and so on). I am then likely to suppose that terrestrial gravity, which is not confined to the equatorial regions, but which is an approximately spherically, rather than axially, symmetric phenomenon, must have an essentially different explanation from the orbits of the moon and planets. Now of course my belief in this theory affects the conditional probability $p(e_2, e_1)$ that I attach to the e_2 that Hesse has derived from the rival Newtonian theory. It is not true that 'its value cannot depend on whether or not we are interested in' Newton's or Huygens's theory: it depends very much on which of these we attach the greatest credence to. There is no question of our assigning degrees of belief to particular predictions of this kind independently of what we think of the various relevant theories available.

To restore any plausibility to her argument, Hesse must make the peculiar psychological assumption that we normally ***first*** assign values to ***all*** the probability functions whose arguments are ***evidential statements*** alone and then allow these values to constrain the values we ***subsequently*** assign to all the probability functions which contain ***hypotheses*** among their arguments. Do we do this? Not only, as the example shows, do we not, but, in fact, it would seem quite impracticable for any realisable, *i.e.* finite, predicting device directly to assign probabilities to all possible combinations of any reasonably large number of possible future individual events. It could only, it seems, generate these (individually, on request) by appeal to general assertions about the interrelatedness of events, that is to say to hypotheses. So not only does Hesse's argument for the irrelevance of theories contain an implicit fallacy, but the relevance of theories seems to be no mere accident of human psychology.

7 *Hesse's Unwillingness to Assign Finite Probabilities to Universal Generalisations.*

Hesse offers the following considerations on pages 109 f.:

> Many philosophers argue that scientific laws and theories are in their essence universal in scope over infinite domains. We shall later find reason to question this, or at least to question its relevance to the logic of confirmation, but if it is accepted, then degrees of belief in laws cannot be represented by betting rates in a game in which there is a fair chance of settling the bet. Nor could the rates in any case be strictly coherent. For if we assume that a law can be falsified but never shown to be true, any bet upon its truth may either be lost or may not be settled, but it can never be won. The only odds on its truth a reasonable man would freely choose under these circumstances would surely be 0:1, for since in any case he cannot win anything, any positive sum he places on its truth may be lost, and at best he will simply not have to pay. A zero degree of belief would thereby be assigned to all laws alike, that is, the same degree that is assigned to statements known to be false.

After presenting independent considerations which show that it is reasonable to regard it as very unlikely that any of our existing fundamental laws should turn out, in the long run, to be exact, Hesse concludes (p. 193) that, 'it is reasonable to have zero degree of belief in any universal generalisation in an infinite domain, as this is actually expressed at any given stage of science'.

8 *Good Reasons for Assigning Finite Probabilities to some Universal Generalisations.*

Hesse is aware of the paradoxical nature of her conclusion. On page 184 she baulks at the idea that any purely existential generalisation must be *certain*, and denies that a probability of unity yields certainty. But later (p. 290) she seems to contradict herself, for there she maintains that all universal laws are *certain* to be false (she herself italicises the word *certain* in both places). Furthermore she also asserts (on p. 185) that 'the use of ideal geometric models in physical theory makes some precise mathematical laws significant, and incidentally gives them non-zero probability'. Her example is the inverse square law of gravitation. Ought she not then to have rejected her original argument as too facile?

If, unlike Hesse, we reject the idea of a bet on an undecidable proposition (as already an absurdity) then we have three choices. We either refrain altogether from assigning probabilities to universal generalisations over infinite domains (except where the proposition in question is decidable: 'Nixon will not re-assume the Presidency of the United States within the next decade' is a decidable proposition which nevertheless apparently generalises over a non-denumerable infinity of temporal moments). Or we extend our probability measure from decidable propositions to universal generalisations over them by appeal to some criterion other than that of coherent bets. Or we extend our concept of decidable propositions by an appropriate mathematical/theological/philosophical fiction to cover the cases at issue and extend the notion of strictly coherent bets to cover the case of wagers with the Devil or other suitable imaginary transactions.

As an example of the second approach, mathematical probabilists would normally require that where $(x)Fx$ is a universal generalisation over a denumerable domain, $p((x)Fx)$ should agree with the limit as n tends to ∞ of $p(F1 . F2 . F3 \ldots Fn)$. This need not be construed merely as a matter of mathematical convenience, but can also be thought of as imposing a kind of ω-consistency on a person's probabilistic beliefs.

As an example of the third approach, let us suppose that Mary Hesse is fairly confident that the comet Kahoutek will return at least once during the next century (say $p = 2/3$), even more confident that if it does so it will do so again the following century, and so on, but assigns a probability zero to the universal generalisation that it will do this for ever. Suppose in particular that she assigns (as well she might) a probability of $(2^{n+1}-2)/(2^{n+1}-1)$ to its passing within the earth's solar orbit during the century n centuries from now, given that it passes through it in all centuries up till then. The Devil bets her a dollar against her soul and a dollar that it will always do so. Consistently with her beliefs she must accept this bet. The Devil then hedges this bet by betting, at her odds, half a dollar that it won't return this century, a quarter dollar that if it does it won't the following century, an eighth of a dollar that if it does it won't the following one, and so on. Then if the comet ever fails to return the Devil breaks even, whereas if it never fails to return He still breaks even financially but gains Mary Hesse's soul. (Indeed if we accept her odds on the decidable propositions but unlike her are ω-consistent with respect to our beliefs about the universal generalisation, we will conclude that there is no chance of her gaining but a fifty-fifty chance that she lose her soul in the fullness of time.) The moral is that the assignment of zero probabilities to all universal generalisations is not only unnecessary but is liable to carry its own kind of incoherence/irrationality with it. But Hesse's position is not inconsistent in the narrow sense and she is in good philosophical company.

9 *The Propriety of Assigning Finite Probabilities to each of an Open-Ended Infinity of Universal Generalisations over Infinite Domains.*

Even when reluctantly accepting the possibility of assigning finite probabilities to some universal generalisations, Hesse is chary of doing so for an infinite number of generalisations. Twice she even seems to assert that this is mathematically impossible:

> These difficulties concern, first, the possibility of assigning finite initial probabilities to all the universal generalisations that may be confirmed by evidence, for if there are more than a finite number of these, assignment of finite probabilities, however small, will violate the probability axioms (p. 141),

> clearly not all of them can have finite initial probability without violating the axiom that the probability of their total disjunction is 1. Indeed only a finite subset of such hypotheses can be assigned probability greater than

> some fixed number, however small, and the rest of the infinite set must have initial probability zero (p. 174).

However, these seem to be just slips since she gets it right on page 176 and she does later briefly discuss Harold Jeffreys's method of assigning finite initial probabilities to denumerably many hypotheses. Her most cogent objection to this is that 'the restriction to differential equations is inadequate even for physics in its present theoretical state'. (I do not understand Hesse when she refers (p. 179) to the problem of 'how to reduce the potentially non-denumerable set of possible hypotheses in any language more complex than Carnap's to a denumerable set'.)

Curiously enough Hesse never even discusses the obvious suggestion that finite probabilities be assigned to all hypotheses expressible in the language on the basis of a lexicographical enumeration. Perhaps she assumes that Putnam's diagonal argument objection to this proposal was fatal. But then what about the following rule: assign a probability of 1/2 to the first hypothesis (or class of hypotheses) that comes to one's attention, 1/4 to the next, 1/8 to the next, and so on, partitioning the probability within any infinite class of hypotheses in the same manner? This method will always leave some probability to spare for the rag-bag of unthought of hypotheses. This would certainly seem to settle the theoretical propriety of assigning finite probabilities to each of an open-ended infinity of universal generalisations over infinite domains. Of course we would not adopt this method in practice. But those of us who apply Bayesian methods in analysing actual examples of scientific inference have in practice no difficulty in assigning finite probabilities to all the relevant hypotheses.

10 *Need we Adopt the Bayesian Prescription for Changing our Beliefs?*

Hesse does give yet another argument against our assigning finite probabilities to a denumerable infinity of theories (pp. 115 ff.), an argument which assumes that the Bayesian prescription for changing our probabilistic beliefs in the light of evidence is not *de rigeur*. Savage proved a certain convergence of opinion theorem which Hesse would like to use in order to prove transitivity of confirmation in certain limiting cases (if h entails e_2, $p(e_2)$ cannot be less than $p(h)$, so we could always raise $p(e_2)$ by raising $p(h)$ sufficiently: convergence of opinion might achieve this). Savage's proof takes for granted the Bayesian prescription for revising probabilities in the light of evidence. Hesse points out that this assumption is not essential to the proof provided 'the initial probabilities . . . have some fixed finite lower limit'. Hesse's formal argument is correct. But the constraint of a fixed finite lower limit has to apply not merely to the actual initial probabilities but to the revised initial probabilities that the non-Bayesian is in some sense committed to by not accepting the constraint on his revised beliefs that the Bayesian prescription requires. Hesse's condition thus imposes a rather severe restriction on non-Bayesian strategies (as well as apparently requiring the estimation of counterfactual and not merely actual degrees of belief) and is not quite as liberating as at first appears. (Hesse's separate argument that Savage's convergence theorem would be no use if there were no lower bound on initial probabilities (since nothing could be proved about the amount of convergence in any finite time) does not seem to me entirely cogent. Of course the rate of convergence depends on the initial probability of the true hypothesis in Savage's scheme: But Hesse, in assigning finite probabilities to only a finite number of hypotheses, may run the risk of not catching the true hypothesis in her net at all.)

In wishing to weaken the conditions of Savage's convergence theorem in order to embrace the possibility of non-Bayesian prescriptions for modifying beliefs in the light of evidence, Hesse follows Hacking and others in assuming that the Dutch Book argument cannot provide a straightforward justification of the Bayesian prescription. This assumption is mistaken and is refuted by a Dutch Book argument attributed to David Lewis and Hilary Putnam (but so simple as hardly to need any attribution) and reported in Teller [1973] (esp. pp. 222–4). (Teller plays down this result (perhaps in deference to his Cambridge teachers or in justification of some of his own further researches) but his reservations about it seem quite insubstantial.) Given the wide currency that has been given to the alleged irrelevance of Dutch Book arguments to the Bayesian prescription one hopes that their demonstrated relevance will now obtain equally wide currency.

11 *The Applicability of Convergence-of-Opinion Theorems to the Case of Actual Scientific Inference.*

Hesse attaches a great deal of weight to the premises required by Savage's convergence-of-opinion theorem. She calls it 'the convergence theorem' and takes it (p. 116) as specifying *necessary* conditions for convergence of opinion (which include assumptions of randomness and independence of tests) rather than merely sufficient conditions. That this is not just a verbal slip is clear from her subsequent discussion which ends with the conclusion (p. 119) that 'convergence of opinion . . . seeks to provide, over and above coherence, *minimal* constraints that are sufficient to yield inductive inference, but this method turns out to be inapplicable in most kinds of scientific reasoning' (my italics). She gets this conclusion by taking these 'minimal' constraints to include randomness and independence of sampling of evidential statements. But it seems to me more correct to regard this theorem of Savage's as illustrative and I think I can prove other convergence-of-opinion theorems which are directly applicable to typical cases of scientific inference.

Let h and g be rival hypotheses and let the evidence *on average* favour h against g, but not to an increasingly negligible extent. Formally, let $\{e_n\}$ be the sequence of evidential statements and let there be some ϵ such that, for any sufficiently large n, $\dfrac{\sum_n p(e_n, h \cdot e_{n-1} \ldots e_0)/p(e_n, g \cdot e_{n-1} \ldots e_0)}{n}$ is greater than $1 + \epsilon$. (This is intended simply to be an accurate formal rendering of the condition I have stated informally in the previous sentence.) I assume that however much the rival theorists disagree about the prior probabilities to be assigned to their respective hypotheses, they can agree, at any rate qualitatively about the ratios in this sum, that is to say that they can agree about what evidence would count significantly in favour of one hypothesis and against the other. I assume that the prior probabilities of both h and g are non zero. I make no assumptions about randomness of evidential tests or independence of different

evidential statements. I claim that these assumptions are already sufficient to guarantee convergence of opinion, i.e. that as the evidence accumulates the ratio of the posterior probabilities of *h* and *g* will eventually become higher than any number one cares to choose. By an elementary theorem in probability, $p(h, e_n . e_{n-1} \ldots e_0)/p(g, e_n . e_{n-1} \ldots e_0) = (p(e_n, h . e_{n-1} \ldots e_0)/p(e_n, g . e_{n-1} \ldots e_0)) \times (p(h, e_{n-1} \ldots e_0)/p(g, e_{n-1} \ldots e_0))$.
Iterating this result, we obtain

$$p(h, e_n . e_{n-1} \ldots e_0)/p(g, e_n . e_{n-1} \ldots e_0) = P_n . p(h, e_0)/p(g, e_0),$$

where P_n is a product of *n* factors of the form $p(e_i, h . e_{i-1} \ldots e_0)/p(e_i, g . e_{i-1} \ldots e_0)$ with *i* ranging from 1 to *n*. Now since, by my earlier assumption, for any sufficiently large *n*, the sum of these *n* factors is greater than $n(1+\epsilon)$, it follows that, for any sufficiently large *n*, their product, P_n, must be greater than $(1+\epsilon)^n$. By choosing *n* large enough, we can make this factor as large as we like, and, since the left-hand side of the equation is the ratio of the posterior probabilities of *h* and *g* and the right-hand side this factor times the ratio of their priors, the theorem follows. In proving this theorem I have explicitly avoided any assumption of independence of evidential statements and I have avoided any assumption of randomness of observational tests by choosing the weaker assumption that the accumulating evidence does on average favour one hypothesis rather than the other. Provided one allows for the fact that in the typical scientific case all the probability assignments will be qualitative rather than strictly numerical, this theorem does seem to me directly relevant to the case of actual scientific inference. So, contrary to Hesse, there do seem to be Bayesian convergence-of-opinion theorems directly relevant to the case of scientific inference. It would be rather astonishing if there were not.

12 *Hesse's Claim to be a Bayesian Personalist.*

Hesse twice claims (pp. 6 and 105) to be a Bayesian *personalist*. Her position seems to me however to be much closer to Carnap's position than to fully-fledged personalism.

She rejects Carnap's system only (*i*) in imposing additional clustering postulates, (*ii*) because of mathematical difficulties which prevent her developing a formal extension of Carnap's system which incorporates her clustering postulates, (*iii*) in not wishing to claim a strictly *a priori* status for her clustering postulates, and (*iv*) in supposing that her basic set of predicates will change with changes in scientific language. These divergences seem to me minor.

Hesse is in significant agreement with Carnap (*i*) in imposing additional constraints on rational belief to those allowed by personalists: 'Rational decision procedures are not sufficient to capture any interesting *inductive* sense of rationality, for which further constraints must be found' (p. 115), (*ii*) in a peculiar obsessional preoccupation with 'basic predicates' which is altogether foreign to the personalist approach: 'On the question of the basic predicates of the language of confirmation theory, Shimony is on strong ground, for all confirmation theorists would agree that some set of predicates must be assumed before the conditions of confirmation can be expressed, and few would object to taking these predicates to be the ones 'seriously proposed' in current language' (p. 178) (in the article referred to, Shimony, a 'tempered', *i.e.* partly-fledged, personalist, had never in fact mentioned *predicates*, let alone *basic* ones, but only 'seriously proposed *hypotheses*').

It would therefore seem to the reviewer less misleading to describe Hesse's position as neo-Carnapian, than to describe it as personalist. Every criticism of Hesse in this review has in fact been merely a defence of simple straightforward personalism.

I have restricted myself here to what seemed to me to be the most central and most questionable doctrines in Hesse's book. There is much else of originality in the book that I have not even touched on. Hesse has interesting things to say about the observational/theoretical distinction, about grue, about ravens, about Maxwell, and about Einstein. Those sections, occupying fully half the book, are virtually unaffected by my attempted critique.

JON DORLING
Chelsea College, University of London

REFERENCE

TELLER, P. [1973]: 'Conditionalization and Observation', *Synthese*, 26, pp. 218–58.

Reviews

MACKIE, J. L. [1973]: *Truth, Probability and Paradox: Essays in Philosophical Logic.* Oxford: Clarendon Press. £4.50. Pp. 305.

The strength of this book is also its weakness: the formidable number of topics treated within two covers. Truth, probability and paradox rate a longish chapter apiece (47, 83 and 59 pages, respectively). In addition there is a chapter each on analysis, conditionals, and dispositions (16, 56 and 34 pages). Links there are among them, occasionally limned, but 'each chapter is a fairly separate study of a topic in philosophical logic' (p. viii). Professor Mackie does preface the book with a pair of unifying ideas. One is that 'we need to take precautions against being fooled . . . by the words we employ'. The other is that we can do so with respect to these topics, 'by the coherent application of fairly simple, common-sense, perhaps old-fashioned ways of thinking' (p. viii). Nevertheless, few readers will persevere from preface to appendix, concentrating instead on a chapter of special interest.

Therein lies the great risk for a work of this sort. The reader with a special interest is likely to be put off by the incompleteness of a single chapter on his subject; unfairly, no doubt, in light of all the rest the book includes, but put off nonetheless. The greater the distance between his views and the author's, the more damaging will the omissions appear. Thus suppose a Quinean were to read, in the chapter on truth, the opening section on the bearers of truth. According to Mackie they are statements (propositions). To the objection that statements are ontologically suspect, or are obscure entities, he replies (in part), 'as a rule we can identify them well enough' (p. 20). But even if this succeeds against certain ontological scruples, it omits the fact that for Quine there is a much more pressing objection to propositions: they would induce determinacy of translation. Mackie nowhere mentions the indeterminacy thesis, let alone the growing literature on its potential implications not just for propositions but for reference and truth. Since he seems repeatedly to be looking over his shoulder at Quine (cf. the Index), even anti-Quineans may wonder at the omission.

Naturally other philosophers will be more concerned about other matters; for instance, the paradox of analysis. Mackie's solution is to explain ways in which an analysis, though expressed in the *form* of an identity, could still be informative. Each way has the consequence that whatever its form, the analysis is not after all a statement of identity between concepts or propositions, and not a statement of necessary equivalence, but something else. But to some philosophers this would seem less a solution than a dodge, since they view the paradox as part of the larger problem of the conditions in which substitutivity of *identity* holds (whether for propositions or anything else), or as requiring distinctions among relevant senses of 'necessary equivalence', or senses of 'synonymous'. None of these alternatives is even mentioned, and none of the relevant literature (Carnap, Church, Pap, *et al.*). Doubtless Mackie felt he had to get on with subsequent chapters. Yet exploring these alternatives would lead to deep problems about the nature of propositions and of synonymy. Even philosophers for whom the paradox of analysis does not arise (*e.g.* Quine) may therefore wonder at the omissions.

Having settled on statements as the truth bearers, Mackie presents his view that the truth of a statement is simply 'for things to be as they are stated to be'. More fully, 'to say that a certain statement S is true is to say that, for whatever *p* we can identify S as the statement that *p*, *p*' (p. 22). What chiefly recommends this view to Mackie, in addition to its extreme simplicity, is that unlike other theories it represents a 'correct conceptual analysis' (p. 40) of 'the central ordinary sense of "true" ' (p. 57). Coherence and pragmatist theories are quickly dispatched as irrelevant to this sort of enterprise. Russellian correspondence theories are dismissed largely for standard reasons.

Tarski's semantic conception receives much more attention. Even so, the account leaves something to be desired. We are told that Tarski's 'aim is ... to provide a criterion for the truth of sentences rather than an account of what "true" *means* even as applied to sentences' (p. 30). And on the following page Mackie implies that Tarski presents 'a set of rules for determining the truth of sentences that shall be mechanically applicable without restriction'. But of course what Tarski provides is *not* a criterion or decision procedure for determining *whether* a given sentence is true, but a way of *defining* (the meaning of) truth for complex sentences in terms of the truth or satisfaction of the simplest clauses of the language.

Something philosophically more serious is neglected when Mackie weighs the relative merits of Tarski's 'very complex procedure' as opposed to his own 'rather trivial one'. He says that 'what extra a Tarskian truth definition supplies is an account of ... the contribution which the internal structure of complex sentences of the language makes to their meaning' (p. 34). He agrees that 'this is additional information of great interest', but goes on to argue that nevertheless his own view ultimately is to be preferred because, for various reasons, Tarski's does not provide a correct conceptual analysis of the central ordinary notion of truth. Instead Tarski provides a 'factual analysis' of a relation between a language and the world which holds when and only when a sentence is true-in-the-ordinary-sense, so that Tarski's term 'true' is at best only co-extensive with the ordinary term (pp. 38, 42).

What is neglected here, among other things, is an important problem Tarski aimed to solve: can semantic notions, including but not limited to truth and reference, be incorporated into a physicalist scientific scheme? (cf. Field [1972]). More generally, are such notions explicable in *any* non-semantic terms? If not, then as Tarski writes in a paper not cited by Mackie, and too rarely cited by others,

> ... it would be difficult to bring [semantics] into harmony with the postulate of the unity of science and of physicalism (since the concepts of semantics would be neither logical nor physical concepts) (Tarski [1956], p. 406).

That Tarski's truth definition shows how the internal structure of complex sentences contributes to their meaning is but a consequence, however important, of his pursuit of the deeper philosophical problem concerning physicalism and the unity of science.

In neglecting this problem, Mackie not only gives a misleading impression of the import of Tarski's theory, he exaggerates the relative merits of his own.

For even if Tarski's programme should prove not to lend itself to physicalist explication of semantic notions, Mackie fails even to inquire whether it is important to try. Indeed he seems to make a virtue of the fact that the notion of a statement, which enters crucially into his definition of truth, is just one of a family of notions such as assertion, command, question, and their contents (p. 52). But while he makes a plausible case that his definition is not thereby circular, he does not note that the family of notions presupposed in his definition, unless physicalistically explicable, is at odds with the fundamental aim of Tarski's enterprise.

With regard to this sort of enterprise, Mackie's analysis of the ordinary notion of truth, even if correct, would be utterly beside the point. What we would want to know is not whether a statement's being true is simply for things to be as they are stated to be, but whether, say, the notion of as-things-are-stated-to-be must be regarded as a non-physical primitive, in the way vitalists regard certain biological notions as physicalistically irreducible, or Cartesians certain mental notions (cf. Field [1972]). Similarly, with respect to statements or propositions, what we would want to know is not so much an answer to the ontological question of whether they exist, but how their existence could be incorporated into a physicalist conception of the world. In general, we would want to know whether there are irreducible semantic facts of any kind.

The chapter on conditionals contains a valuable survey of different kinds of conditionals and related forms. The survey is undertaken to establish what range of cases an adequate account would have to cover, and to find 'a unitary theory of the core meaning of if-sentences' (p. 104). But why must there be a single meaning of if-sentences in the first place? It would seem more plausible to argue that there are several, as he does later (Chapter 5) for concepts of probability, related by family resemblances. Six rival accounts of conditionals are considered and found wanting, usually but not only for failure to cover all the cases revealed by Mackie's survey. The six are the materialist, the logical-powers, the consequentialist, the metalinguistic, the telescoped-argument, and the possible-worlds accounts. His treatment of the possible-worlds accounts is not as strong as his treatment of the others. He objects that talk of possible worlds calls out for further analysis; that anyway there *are* no possible worlds except the actual one; that such an account proposes a linguistic reform instead of an analysis of what is ordinarily meant; and that it fails to give truth conditions for counterfactuals, but only acceptability conditions (pp. 88–92).

Not one of these objections is of high quality. The last is just false; one need only consult David Lewis's definition, quoted by Mackie himself (p. 88), to see that the definition is in terms of truth conditions. (Also cf. Lewis [1973], pp. 18–20 and index.) Mackie's contrary interpretation depends on his loaded way of construing possible worlds and their similarity. With regard to the first objection, doubtless possible worlds call for further analysis, but the crux of Mackie's own proposal calls for even more (the concept of asserting something within the scope of a supposition, about which Mackie himself expresses misgivings on pp. 98, 103, and which I shall argue is circular). As to whether there are non-actual, possible worlds, of course in one sense there aren't: only our world is actual. But surely we do not want to deny that there are other ways our world might have been. This, so far as 'existence' is concerned, is all that most possible-world theorists really need. Even Quine ([1969], pp. 151–2) makes room for possible worlds, as different distributions of particles throughout space-time. Finally, that what is proposed is a linguistic reform—if it is—might be a good thing, since the ordinary concept of a conditional—if there is just one—may be problematic, especially if Mackie's account of it is correct.

At the heart of Mackie's account is this general analysis:

> (C) To say 'If P, Q' is to assert Q within the scope of the supposition that P (p. 93).

Most of the objections that are likely to occur to anyone are considered by Mackie, often deftly. But there is room for further consideration of the objection that (C) is either circular or contains an analysans not substantially clearer than the analysandum. It may look circular, as Mackie notes, because to assert Q within the scope of the supposition that P seems just to assert Q on condition that P, which seems just to say 'If P, Q' (p. 98). Mackie responds,

> To meet this criticism we must explain supposition without reducing it to saying 'if'. And we can. Philosophers in particular must be familiar with the procedure of starting off with 'Let us suppose, just for the sake of argument, that . . .' . . . and going on to develop the consequences of this suggestion. . . . The use of 'if' can be understood in light of this sort of procedure; this sort of procedure does not need to be understood in terms of 'if' (p. 98).

Of course we are familiar with this procedure. But we are, or ought to be, equally familiar with the problem of explaining what is meant by developing the *consequences*. 'Will not conditionals crop up in the explanation of the specific types of consequence?' Ironically it is Mackie himself who asks this question, in another context (p. 83). It constitutes the crux of his objection that the 'consequentialist account' of conditionals is likely to be circular. If the objection applies there, why not here?

Let us suppose, just for the sake of argument, that 'if' need not crop up in an explanation of the concept (or concepts) of consequence. This would free (C) of circularity. Still, the notion of consequence has to be understood *somehow*, and there seems to be no more agreement on how to understand it (or how to understand certain crucial notions of consequence) than on how to understand conditionals. Therefore the analysans in (C) seems no clearer than the analysandum, if asserting Q within the scope of the supposition that P just amounts to developing the consequences of the supposition that P. If Mackie replies that what he has in mind is the human *activity* of developing or envisaging the consequences (pp. 98, 100), or of inferring (p. 93), instead of the *concept* of consequence, I still fail to see any gain in clarity. How could we identify the relevant activity or behaviour, save in terms of ascribing to the agent something called 'inferring' or 'envisaging' or 'developing' the *consequences*? All the troubles with 'consequences' would be visited on the notion of the activity of developing the consequences.

With regard to Mackie's explanation of the type of consequence involved in counterfactuals, much needs to be said. I suspect others will soon be saying it, so I limit myself to some remarks about one of Mackie's two main arguments that counterfactuals (like all conditionals in their basic use) are neither true nor false (p. 105). (The other argument fails too, but I omit it here.) A basic premise in the argument is that 'simple truth, truth in the strict sense, can belong only to descriptions of what is actual' (p. 106). This seems extremely implausible. On one interpretation, descriptions of the past would be neither true nor false; so too for descriptions of the future, if we could call them that. On the other interpretation, in which what is actual includes what actually was the past and what actually will be the future, it precludes the truth even of a law of logic applied, say, to any distribution *D* of particles throughout space-time which differs from the actual distribution. We could not assert that *D* either contains stars or not; nor could we assert that *D* is preferable to the actual rather unfortunate distribution.

Mackie's claim that truth (or falsity) can belong only to descriptions of what is actual is supposed to follow from his definition of simple truth: as things are stated to be, so they are. Here is the argument verbatim:

> 'In the possible situation that *P*, *Q* also', just because it purports to describe a merely possible situation, does not leave room for things to be, or not to be, as is stated (p. 106).

Even if this were a good argument, we should simply conclude, 'So much the worse for Mackie's notion of truth', in view of its inapplicability even to (extensional) examples like non-actual distributions of particles. But the argument is a non-sequitur. If truth is for things to be as they are stated, then of course it is for *things actually to be* as they are stated. But it does not follow that truth is for *actual things to be* as stated. Nor does anything in Mackie's earlier discussion of statements imply that an utterance of a sentence expresses a statement only if what it purports to describe is actual (cf. pp. 18–22). One who utters 'In the possible situation that *P*, *Q* also' normally succeeds in making a statement. How things are stated to be by such a person is simply that in *P*, *Q* also; and things are as they are so stated if and only if in *P*, *Q* also. *P*'s being a non-actual situation is irrelevant to whether there is room for things to be as stated, unless Mackie is using an odd notion of there being room for things to be as stated, for which his earlier discussion of statements and truth provides no basis.

In the course of the chapter on dispositions, Mackie presents a brief consideration against construing the meaning of 'believes that . . .' along dispositional lines. In doing so he makes an extraordinary claim:

> Despite all the criticisms made in recent philosophy of the notions of introspection, private languages, and the like, it is a hard fact that we know more about the state of believing, as it is in itself, than we do about the states that form the temperature-scale as they are in themselves . . . (p. 147).

It seems slightly preposterous to call this a *hard fact*. But what is more disturbing is the strong hint of unshakable belief in the adequacy of armchair philosophising. This goes rather beyond Mackie's reasonable remarks in the preface on behalf of simplicity and common sense. It suggests a touching faith in our manifest image of the world, as it is enshrined in ordinary language. One wonders how seriously to take Mackie's disavowal of dogma about our ordinary conceptual scheme's being in perfect order (pp. 9–10).

By far the longest chapter of the book is the one on probability. Mackie thoroughly characterizes five concepts of probability, and argues that all five can 'be coherently developed and made at least reasonably clear and precise'; that no one (or even two) of the concepts covers all the uses of probability language; that nevertheless there are family resemblances growing out of a central root by a series of natural shifts and extensions; and that because of this, one needs 'not merely to make sure that one does not slide from one sense of "probable" . . . to another [but] to decide whether a particular transition is one of the justifiable ones or not' (p. 156). A substantial part of the chapter is therefore devoted to analysing these meaning-shifts and sorting out valid from invalid transitions in arguments about probabilities. It is from this point of view that Mackie investigates how to interpret and apply Bernouilli's Central Limit Theorem and Bayes's First and Second Theorems.

Obviously philosophers of probability will find much to use here, whether as ammunition or as a target. But many other philosophers will wonder whether all the clarificatory work, all the brush clearing, is worthwhile. Not everyone holds, with Mackie, that

> if there are good but not conclusive reasons for believing things, there must also be relations of quasi-entailment, of partial, non-deductive, support (p. 166). . . . To be rationally justified a set of beliefs must be such as to lend itself to measurement by a coherent set of simple probability values (p. 231).

On the contrary, good reasons could consist of survival of criticism, or coherence with the simplest over-all conceptual scheme, or whatever, rather than quasi-entailment from data to theory. Those philosophers who hold that rationality, whether in science or elsewhere, does not consist in *justification* but either in criticism, or in the pursuit of research programmes, or in inference to the best total explanation, will not rate exacting analysis of probability very high on their list of priorities; similarly for those who hold to justification but question whether it is susceptible to measure, even in rough terms, especially between two theories (or perhaps two 'paradigms') divided by a scientific revolution. Thus the chapter will seem inadequately motivated from the point of view of many readers, however valuable it may be as an exercise in taking precautions against being fooled by certain words we employ. For what if we need not employ the words?

Mackie's motive for his final chapter is primarily to test the capacity of his simple theory of truth to solve the paradoxes. I applaud the motive; also the wide variety of paradoxes he attempts to solve in a unitary way. Unfortunately, it may not be wide enough. For example, he has neglected 'grounding paradoxes' (Herzberger [1970]). The oversight has adverse consequences for Mackie's treatment of the 'logical proof approach', exemplified by J. F. Thomson but much more richly developed by Herzberger, whom Mackie nowhere cites. Mackie's criticism of the approach is based on his unargued assertion that the sense of the phrase 'not truly applicable to itself', standardly construed, 'is *there*, in the English language, already, and no doubt has equivalents in most other languages' (p. 256). But if Herzberger is right, *no* language can express such a concept, contrary to armchair appearances; indeed no language can express its grounding concepts. Mackie also neglects paradoxes which involve modality, such as the Possible Liar ('What I say is *possibly* false'). Variations of this paradox cause trouble for solutions which are uncomfortably similar to Mackie's own (*e.g.* Kneale, Ryle, Sommers; cf. my [1973], [1974]).

Mackie considers—and rejects—a wide variety of competing solutions. But again the variety may not be wide enough. I have mentioned Herzberger's solution. There are others, perhaps the most influential of which is Bas van Fraassen's [1968]. In addition, a really complete treatment would include Brian Skyrms [1970], Robert L. Martin [1967], Zellig Harris [1968], and Fred Sommers [1969], to mention only some who had published in time to have come to Mackie's attention. In every case, failure to take account of this literature has adverse effects on Mackie's presentation. For example, he asserts of the Liar Paradox that 'there is not even anything like a category mistake' involved (p. 242). But Martin argues in a series of papers that there *is*, and that this provides the key to a solution (cf. my [1973] for references). Even if Martin is proved wrong, his work should warn us against asserting so confidently that the Liar involves nothing like a category mistake. (Mackie's qualifying remarks on pp. 243–4 are beside the point Martin makes.) Similarly, the work of Harris and Sommers, even if ultimately inadequate, should warn us (in different ways) against armchair assertions that Liar-like sentences are perfectly grammatical (p. 242).

With regard to Mackie's own solution, I shall comment on just one of its troublesome aspects. As he notes, his solution implies a sort of type distinction: 'Higher order properties occur only where they arise from some first-order properties or states of affairs' (p. 268). He is concerned to 'prevent the construction of circles that would result in (direct or inverse) self-dependence' (p. 254), and to justify the principle that 'nothing can really depend on itself' (p. 253). His justification is that the principle has an independent philosophical ground in his simple notion of truth; it is not introduced *ad hoc* simply to exclude paradox (p. 268). All this is applied to the set-theoretical paradoxes as well as the semantical. Further, we are told that we can incorporate his principle into axiomatic set theory, if we choose, once it has been provided with his extra-systematic philosophical justification (p. 266). Since Mackie's principle so closely resembles the vicious circle principle (pp. 253, 266, 272), one wonders how it should be incorporated into set theory without blocking results we want to keep—*e.g.* least upper bounds, Cantor's Theorem, and Cantor's diagonal proof.

The trouble here is threefold. I present it in order of increasing importance. First, no instructions are given about what axiomatic form Mackie's principle would take. In part I suspect this is because, second, the principle is very vague—even more than Russell's. Mackie should have tried to clarify his principle, as Gödel and others have for Russell's. (Cf. Gödel [1944]; and Chihara [1973], which also contains further references.) Third, Mackie's approach *would* block the diagonal argument, contrary to his efforts to show that he can solve Richard's paradox without impugning the diagonal result (pp. 274–5; Mackie nowhere deals with upper bounds or Cantor's Theorem). We are told that the Richard definition is based on itself, and 'incomplete', in the sense that it would actually be in its own domain of reference if and only if it 'names' a real number (p. 273); its being a successful definition turns out to depend on its being a successful definition. This 'incompleteness' is supposed to be radically different from anything in Cantor's procedure:

> Cantor's ***reductio ad absurdum*** starts with the sheer ***supposition*** that all the non-terminating decimals are already there as an ordered set. . . . There is no similar incompleteness in Cantor's definition of *d* [the diagonal decimal], since it is not (as Richard's would be if its 'definition' were a definition) based on itself; Cantor's *d* is based on the ***supposedly*** already ordered set of non-terminating decimals. (We cannot argue that there is a similar incompleteness here on the ground that *if* the decimals formed an ordered set, whichever one turned out to be *d* would be there already, say as the *r*th decimal in the table, and therefore that the definition of *d* would include the incoherent instruction to make its *r*th decimal figure something different from what it is; the proof shows that there cannot be a decimal in the table which may turn out to be *d*. . . .) (Pp. 274–5; Mackie's emphasis.)

'Supposition' is a red herring here (as is 'ordering', in the unquoted remarks on p. 275). True, the proof *does* show that *d* is not in the table. But one of the *steps* in the proof is a definition of *d* as the decimal whose *n*th figure is 1 if the *n*th figure in the *n*th decimal is 2, and 2 otherwise (p. 269). This definition occurs within the scope of the *reductio* supposition that *all* the decimals in the interval between 0 and 1 have been paired with the integers. The *definition* of *d* therefore proceeds in terms of the supposition that *d is* already among the decimals in the interval. In terms of this supposition, the definition of *d* is based on a totality to which *d* is supposed to belong (cf. p. 272). Under the *reductio* supposition, therefore, *d*'s being a decimal in the interval appears to depend on its being a decimal in the interval, thus violating Mackie's principle that nothing can depend on itself. Worse, under the supposition that all the decimals in the interval have been paired with the integers, *d would* be the *r*th decimal, say, and thus, contrary to Mackie, the definition of *d would* include the instruction to make its *r*th decimal figure something different from what it is. But Mackie holds this sort of instruction to be incoherent (see also p. 271).

I conclude that the diagonal result *is* blocked on Mackie's approach (and that the same is likely to hold for least upper bounds and Cantor's Theorem). In light of the similarity of his approach to the vicious circle principle, or to related efforts to block 'impredicative' procedures, it is not surprising that he has trouble here. What is surprising is how superficially he appears to have dealt with it, given the detailed discussions in the literature since the problem first bedevilled Russell. One suspects this is an area where Professor Mackie's faith in 'simple, common sense, perhaps old-fashioned ways of thinking' has led him astray. This is unfortunate, since his book is a *tour de richesse*, from which philosophers can glean much of value that I have had to omit altogether, in the interests of space and the spirit of loyal criticism.

JOHN F. POST
Vanderbilt University

REFERENCES

CHIHARA, C. [1973]: *Ontology and the Vicious Circle Principle.*
FIELD, H. [1972]: 'Tarski's Theory of Truth', *Journal of Philosophy*, 69.

FRAASSEN, B. VAN [1968]: 'Presupposition, Implication, and Self-Reference', *Journal of Philosophy*, **65**.

GÖDEL, K. [1944]: 'Russell's Mathematical Logic', in P. Schilpp (*ed.*): *The Philosophy of Bertrand Russell.*

HARRIS, Z. [1968]: *Mathematical Structures of Language.*

HERZBERGER, H. [1970]: 'Paradoxes of Grounding in Semantics', *Journal of Philosophy*, **67**.

LEWIS, D. [1973]: *Counterfactuals.*

MARTIN, R. L. [1967]: 'Toward a Solution to the Liar Paradox', *Philosophical Review*, **76**.

POST, J. F. [1973]: 'Shades of the Liar', *Journal of Philosophical Logic*, **3**, pp. 370–86.

POST, J. F. [1974]: 'Propositions, Possible Languages and the Liar's Revenge', *British Journal for the Philosophy of Science*, **25**, pp. 223–234.

QUINE, W. V. [1969]: *Ontological Relativity and Other Essays.*

SKYRMS, B. [1970]: 'Return of the Liar: Three-Valued Logic and the Concept of Truth', *American Philosophical Quarterly*, **7**.

SOMMERS, F. [1969]: 'On Concepts of Truth in Natural Languages', *Review of Metaphysics*, **23**.

TARSKI, A. [1956]: *Logic, Semantics and Metamathematics.*

SHEA, W. R. [1972]: *Galileo's Intellectual Revolution.* London: Macmillan. £4.95. Pp. xii + 204.

Shea has written an important and useful book. His book covers the neglected middle period of Galileo's work and concentrates almost exclusively upon the period between Galileo's first astronomical discoveries and the publication of *Dialogue Concerning the Two Chief World Systems.* The book opens with a summary and analysis of the *Discourse on Floating Bodies* and its Archimedean background, proceeds to discuss the *Letters on the Sunspots* and the controversy of the comets, and ends with three chapters presenting and laying out the arguments of the *Dialogue.* Shea is careful throughout to present Galileo in the context of the contemporary controversies which raged around his work. Of greater interest to the readers of the *Journal* is the fact that Shea also includes in most chapters a section discussing the methodology behind Galileo's work.

Somewhat regretfully, Shea does not discuss the episodes and controversies surrounding the telescope in any detail, though these took up a large part of Galileo's time and effort. Such an omission may be thought unimportant because Galileo's telescopic work does not exemplify his methodology of science. This, however, would be a strange conclusion, since the type of argumentation which is carried on in *Siderius Nuncius* and in the reply to Galileo's critics by Wodderbornius is quite similar to that which we find Galileo using in *Letters on the Sunspots* and the *Dialogue.*

If Shea's book suffers from a single overall defect, it is his too unhesitatingly siding with Koyré in dealing with Galileo's methodological remarks and designs. Galileo is seen primarily as a Platonist using mathematical reasoning to achieve certainty in his work. Experiments are only introduced, if at all, to verify the certain conclusions obtained by this mathematical analysis. While this much seems to be true and to the point, it does not go far enough. Galileo's relation to Platonism is as yet ill-understood. In his works generally, but especially in the formative period represented by the Juvenilia, Galileo makes it clear that he was reading many of the Neo-Platonists, both ancient and modern. He reports in lists the opinions of Plotinus, Proclus and Iamblichus, as well as takes up, in various contexts, ideas related by Telesio, Cardano and Ficino. But most of these references do not seem to deal with the Neo-Platonic conception of certainty through mathematical reasoning (or mathematical causes). The exact relation and debt of Galileo to his Platonic predecessors is left unclear.

Another aspect of this general problem becomes evident when one reflects that Galileo's use of mathematics in *Letters on the Sunspots*, *The Assayer*, and the *Dialogue* is quite simple and not very extensive. Certainly Galileo many times affirms his faith in the necessity of using mathematics and mathematical reasoning to attain conclusions which are certain regarding natural phenomena, but he does this at the expense, in the *Dialogue* for example, of not treating Copernicus' system in all its mathematical complexity (*i.e.* as having numerous epicycles, *etc.*). Perhaps Shea would argue, as he does in discussing Galileo on the tides, that Galileo was quite prepared to sacrifice the details of a particular theory in order to achieve certainty with regard to its basic principles. But in the case of the planets, unlike that of the tides, Galileo knew that if he retained the necessity of circular motion (which his inertial principle as argued for in the First Day requires), he would have to add epicycles, eccentrics or some such mathematical device, in order even to come close to saving the phenomena.

A third point concerning Galileo's methodology appears when one considers the varying and strange remarks Galileo makes concerning causes and essences (in *Siderius Nuncius* and *The Assayer*). It seems that Galileo still recognises and requires causal explanations in the classical sense. One way of reading his Platonism is to require, as some Platonists did, that the formal and final causes involved be stateable as mathematical predicates. To my knowledge, except as briefly touched upon in the work by Charles Schmitt and others concerned with Galileo's relation to the Paduan school, causes are never closely examined in any treatment of Galileo's method. Yet, on both the Aristotelian and Platonic accounts, the role of causes in explanation and proof is paramount.

These comments are not intended as a direct criticism of Shea's book. Rather, they indicate what needs to be done concerning Galileo: specifically, an examination of the kind and character of the Platonism he is using and its background, and, generally, a study of sixteenth-century methodological thought. There have been some recent moves in this direction, but much is left to be done. A full treatment of these issues would involve discussions of causes and the role of mathematics, and, as such, is a necessary prerequisite to our understanding of the scientific revolution. Shea is aware of these problems, as part of his text and footnotes indicate. What one must be thankful for is that we now have a detailed treatment of Galileo's works during a crucial twenty-year period, and that they are treated in such a fashion that the results will be of great use. Furthermore, Shea's book may provide a focus for future discussions of Galileo's methodology and for more general investigations in the methodology of science. Philosophers will never again have the excuse, when characterising Galileo's methodology, that they can 'get all that is needed' from the *Discorsi* and the problem of free fall.

PETER MACHAMER and STEPHEN LUNSFORD
Ohio State University

RUSSELL, Bertrand [1973]: *Essays in Analysis.* (Edited, with introductory notes, by Douglas Lackey.) London: George Allen & Unwin Ltd. £5.50. Pp. 345.

Mr Lackey states his main aim in compiling this anthology to be that of making clear the development of Russell's thought in the period 1904 to 1913. He has selected fifteen of Russell's essays, mainly from these years: three hitherto unpublished, three published only in French, and others of varying degrees of accessibility. He remarks that the establishment of the Russell Archives of McMaster University has inaugurated the era of Russell scholarship, and as an aid to such scholarship the book is to be welcomed. More general justification for the publication of a selection of this sort must be sought in the independent interest of the subject-matter, and of Russell's treatment of it. Mr Lackey, in his introductory comments, attempts to relate the material to current concerns. The links he finds, though, are at times far-fetched, and this up-dating project leads him to mistaken accounts of Russell's development.

The book divides into four sections, differing widely in interest. The first—'Russell's Critique of Meinong'—consists of the three-part series 'Meinong's Theory of Complexes and Assumptions' from *Mind* (1904), and two reviews of books by Meinong and his followers. At this time Russell felt himself to be in substantial agreement with the Graz school, despite noting that Meinong saw himself as a psychologist while Russell regarded his own field of study as logic. The effects of this divergence were minimised by Russell's employment of a Moorean methodology in 'logic'—*i.e.* detailed 'inspection' of the facts or concepts under analysis. Russell's work here is archaic, not because of any ontological extravagance (hardly greater than current acceptance of 'possible worlds') but because it so heavily embodies the methods of the early Moore.

The second group contains three essays on descriptions and existence. In his Preface, Mr Lackey rests his decision to exclude 'Mathematical Logic as based on the Theory of Types' on the fact of its ready availability elsewhere. This argument would seem to apply equally to 'On Denoting' and 'Mr Strawson on Referring' reprinted in this section. The third essay here—'The Existential Import of Propositions'—is a short reply to two papers by the Scottish logician Hugh MacColl, who responded in two further notes. The MacColl papers are included in an appendix. Lackey finds the crux of the debate in the question 'Are there entities which are not existents?'—Russell's answer being 'no', MacColl's 'yes'. He sees in MacColl's position an anticipation of one way of developing presupposition-free logics. In order to make sense of the exchange in these terms, he states that Russell's paper contains 'for the first time the view of existence that became characteristic in his later philosophy'. Crudely, in one particular case, this amounts to the claim that 'Men exist' means no more than 'There are men'. Some might read this in turn as 'The class of men has at least one member'. Russell's paper indeed alludes to propositions such as this latter, but not as an analysis of 'Men exist' but of 'The class of men exists'. What he in fact states is, almost word for word, the 'early' theory enunciated in *The Principles of Mathematics*. In both works we are introduced to a 'technical' sense of 'exists' which means 'is a class which has at least one member'. In both we are informed of another sense of existence which the present paper dismisses as irrelevant to Symbolic Logic. In *this* sense, entities in space and time exist but some entities for example those of mathematics, do not. On either reading of 'exists', the Russell of this paper would maintain that there are entities which are not existents ~ mathematical objects in the second sense, the null class (and indeed anything not a class) in the 'technical' sense. The paper in question was published three months before 'On Denoting', and embodies the rejection found there of *unreal individuals*. But the explicit account of existence comes straight from the *Principles*. No simple theory of an early and late view of existence can do justice to the facts of Russell's development.

Russell's major efforts in this period were devoted to the paradoxes and to the development of the theory of types. Not surprisingly, the section devoted to these matters is the most substantial and retains a contemporary interest largely lacking elsewhere. Gödel, Quine and Chihara, among others, have taken Russell's views as starting points for important discussions in the philosophy of set theory and have of necessity referred to papers here reprinted. These are, under their published titles—'On Some Difficulties in the Theory of Transfinite Numbers and Order Types' of 1905, 'Les Paradoxes de la Logique' and 'La Théorie des Types Logiques' of 1906 and 1910 respectively, published in French but here reproduced from the original English manuscript. Hitherto unpublished is a 1906 paper 'On the Substitutional Theory of Classes and Relations'.

The first paper shows Russell considering merely the set-theoretic paradoxes—the semantic ones are ignored—and outlining the general form of a solution: some propositional functions do not determine classes. The problem is—which do? Three theories are sketched. Two retain classes corresponding to *some* propositional functions—in the first where the functions are in some sense 'simple', in the second if the resulting class would not be 'too large'. In an added note of 1906, Russell opts for the more radical third theory, the 'no-class' theory: we need *never* assume that classes exist.

In *Principia* the 'no-class' theory takes the form of a contextual definition of class-notation in terms of propositional functions. Quine has argued that if we avoid Russell's use-mention confusions, propositional functions are attributes, differing from classes only in their identity conditions. If we are working with a *simple* theory of types (in effect guaranteed by the Axiom of Reducibility) the obscurely strong identity conditions of attributes are pointless and we might as well adopt classes from the start.

For Russell, difference in identity conditions was certainly *one* line of demarcation between propositional functions and classes. Another though, as these early papers bring out, was rather the Fregean contrast between concepts and extensions, the point being that the former are *incomplete*, the latter complete. In terms of symbolism this requires that any expression for a concept contains a gap for its argument. Russell expresses a similar view—'. . . a function, as Frege himself has rightly urged, is nothing at all without some argument . . .'. Commentators on Frege have noted that strict adherence to this notational requirement enforces a doctrine of simple types for concepts, one which cannot be violated in his formal language. This is precisely what Russell attempts to build into his 'substitutional theory' here: formulae in non-primitive notation which violate simple type restrictions contain, when written out in primitive notation, one or more unfilled argument-places. Even by Russell's standards, the exposition in these papers is confused. Nevertheless they are essential to any attempt to understand the many strands which make up the mature theory of types.

The material in the final section (short pieces on the philosophy of logic and mathematics) is on the whole of little interest. The three previously published pieces are either polemically directed at uninteresting targets, or of a popular nature. In an unpublished essay of 1907 Russell states more fully the view, found in *Principia* and elsewhere, that the premises of mathematics, laid bare in the logicist programme, derive their support from their consequences and not *vice versa*. Russell can hardly be said to have done much to develop a philosophy of mathematics on this view, which squares so oddly with the generally Cartesian structure of his epistemology. A final unpublished essay, tentatively dated 1950–2, records Russell's acceptance of the view that the propositions of mathematics and logic are purely linguistic, but makes no significant contribution to its defence.

Mr Lackey includes three quite useful bibliographies: of the historical background to Russell's early logical work, of Russell's writings on logic, published and unpublished, and of secondary material on the topics of the book.

I. G. MCFETRIDGE
Birkbeck College, London

STOVE, D. C. [1973]: *Probability and Hume's Inductive Scepticism*. Oxford: Clarendon Press. £3.00, pp. 132.

David Stove has written a good book. In it he claims to identify and assess Hume's only argument for his inductive scepticism, and to refute both this scepticism and the main controversial premiss in that argument, deductivism. I shall argue that both refutations fail. But the clarity of Stove's bristly, professional prose (and, better, thought) and the care with which he has made explicit, made precise and defended any possibly controversial assumptions of his own set standards rarely attempted. On Humean exegesis, on Hume's influence, and in diagnosing and rebuking misjudgments of these, I find him informative, usually persuasive and sometimes acidly devastating.

His argument against inductive scepticism is tight, and its abridgment risks distortion. It hinges on at least

(*i*) the theorem that when $P(\sim p \vee \sim q, r) < 1$ and $P(q, r) < 1$, $P(p \,\&\, q, q \,\&\, r) > P(p \,\&\, q, r)$. In forms apparently relevant to the problems of induction to, respectively, a generalisation and an unobserved case, the consequent is specialised, respectively, to '$P(\forall x Fx, Fa \,\&\, t) > P(\forall x Fx, t)$' and '$P(Fb \,\&\, Fa, Fa \,\&\, t) > P(Fb \,\&\, Fa, t)$', where '$t$' represents any tautology. This part of the argument is due to Manfred von Thun, and is unassailable.

(*ii*) the interpretation of the principles of the probability calculus as those of logical probability, and the interpretation of statements of logical probability as those of the degree of conclusiveness of the argument from (consistent) premisses to a conclusion. Thus Stove interprets inductive scepticism and deductivism as the claims that for all e and h such that the argument from e to h is, respectively, inductive or invalid, $P(h, e \,\&\, t) = P(h, t)$.

(*iii*) the assumption that the arguments from '*Fa*' to '*Fa* & *Fb*' and from '*Fa*' to '$\forall xFx$' are inductive.

I think that there are two independent and adequate comebacks to this neat argument, the first of which attacks (*iii*).

The sense of 'inductive' required is 'predictive–inductive', if, as I believe, Stove is right about Hume: 'It is not about all inductive inferences, but only predictive–inductive ones, that there exists in Hume an argument which is explicit and clear, and which ends in a sceptical conclusion' (p. 29). Thus it is the argument to '*Fb*' rather than that to '*Fa* & *Fb*' that we should consider. As one reviewer of the book remarked,[1] von Thun's conclusion would follow equally if we replaced '*Fb*' by '$\sim$ *Fb*'. Since '*Fa*' increases the prior probability of '*Fa* & *Fb*' by exactly the same token that it increased that of '*Fa* & $\sim$*Fb*', it is not reasonable to read this argument as supporting the claim that '*Fa*' increases the prior probability of '*Fb*'. So the von Thun–Stove argument fails to refute scepticism with respect to the *predictive*–inductive *component* of the inference from '*Fa*' to '*Fa* & *Fb*'; and so fails, by parity, in the case of the generalisation.

The *TLS* reviewer made his remark in order to refute what he took to be Stove's attempt to solve the problem of induction. Professor John Mackie[2] succinctly corrected him, quoting Stove (p. 2): 'what I claim for the argument is not . . . that it is a justification of induction; but just that it is a refutation of Hume's scepticism about induction.' The reviewer took the point, but deplored a choice of examples which misleadingly suggested relevance to the larger problem. (Hence my more general statement of the theorem, and my use of '*apparently* relevant', above.) I make the point to a different end—against Stove's claim, endorsed by Mackie and conceded by the reviewer, to have refuted Hume's specifically predictive–inductive scepticism.

Still, replacing (*iii*) with the (*iii'*) defined by replacing the 'inductive' of (*iii*) with 'invalid' renders it acceptable, and might, for all I have said so far, leave us at least a disproof of deductivism.

(I thus simplify Stove's main argument against deductivism, which derived the invalidity of these arguments from the false assumption of their predictive–inductive characters. He provides a battery of other arguments not using von Thun's theorem, but each includes premisses whose acceptance rests evidentially on some intuitive assessment of difference in cogency among invalid arguments. Since what he is trying to prove is precisely the existence of such differences, these arguments are question-begging. The rebuttals he offers to this charge manage to miss the obvious with some panache.)

The second reply to Stove attacks (*ii*). I want to say that (*a*) a tautology provides no evidence at all for any contingent claim, and for this reason does not provide better evidence for some such than for others, so 'arguments' from tautologies to contingent conclusions are not more or less cogent, and (*b*) there can be—as any teacher of the subjects can testify—bad, uncogent arguments even for theorems of mathematics and logic. I agree, for good measure, that there are at least two contingent statements, p and q, whose conjunction and conditional are likewise contingent. Stove would consider my position demonstrably inconsistent; it is inconsistent with (*ii*). For Stove will read *a* as entailing that $P(p, t) = P({\sim}p, t) = P(p \,\&\, q, t)$, and, since $P(p, t)+P({\sim}p, t) = 1$, that each

---

[1] *Times Literary Supplement*, 10 August 1973.

[2] *Ibid.* 28 September 1973.

$= \frac{1}{2}$. But since $P(p \,\&\, q, t) = P(p, t) \times P(q, p \,\&\, t)$, I must now admit that $P(q, p \,\&\, t) = 1$, contradicting my assertion that $p \supset q$ is contingent.

Stove does not use this argument, but he uses all the premisses and rules of interpretation it requires; von Thun's more complex arguments rest on a similar trick. The analogous 'refutation' of *b* from (*ii*) is too simple to elaborate. Stove and I would agree that neither *a* nor *b* is consistent with (*ii*); finding the former much more plausible, I reject (*ii*), while finding the latter compelling, he will credit me (I presume) with having refuted the former. So what Stove takes to be a disproof of deductivism, I take merely as a proof that deductivism is also inconsistent with (*ii*), which is independently unacceptable.

JOHN FOX
La Trobe University

BOSTON COLLOQUIUM FOR THE PHILOSOPHY OF SCIENCE

FIFTEENTH ANNUAL PROGRAMME, 1974–75

Conventionalism Revived, Poincaré Vindicated
HANS EKSTEIN, Physics, CNRS, Centre de Physique Theorique (Marseille)
Commentator: PETER HAVAS, Physics, Temple University
Tuesday, 24 September 1974

The Peculiar Evolutionary Strategy of Man
LAWRENCE B. SLOBODKIN, Ecology and Evolution, SUNY, Stony Brook.
Commentator: STEPHEN J. GOULD, Geological Sciences, Harvard University
Tuesday, 8 October 1974

Is There Something Underlying Language Which is Deeper Than Logical Representation?
WALLACE L. CHAFE, Linguistics, University of California, Berkeley.
Commentators: JERROLD J. KATZ, Philosophy, M.I.T. and CUNY, City College.
MATHILDA S. HOLZMAN, Child Study, Tufts University
Tuesday, 15 October, 1974

Concepts of Ontology and Time in Anthropology
ANTHONY LEEDS, Anthropology, Boston University
Commentator: ANNEMARIE SHIMONY, Anthropology, Wellesley College
Tuesday, 22 October, 1974

Logic and Dynamics in a Fuzzy World
MICHAEL A. ARBIB, Computer and Information Science and ERNEST MANES, Mathematics, University of Massachusetts, Amherst
Commentators: DANIEL C. DENNETT, Philosophy, Tufts University and
JOSEPH WEIZENBAUM, Electrical Engineering, M.I.T.
Tuesday, 12 November, 1974

Causal Versus Teleological Explanation of Human Action
RAIMO TUOMELA, Philosophy, University of Helsinki and Princeton University
Commentator: RODERICK M. CHISHOLM, Philosophy, Brown University
Tuesday, 19 November, 1974

Cosmological Implications of Anomalous Redshifts
JEAN PIERRE VIGIER, Theoretical Physics, Institut Henri Poincaré (Paris)
Commentator: JOHN STACHEL, Physics, Boston University
Tuesday, 3 December, 1974

The Compton Effect: Historical-Philosophical Remarks
ROGER H. STUEWER, Physics and History of Science, University of Minnesota and History of Science, Harvard University
Commentator: AAGE PETERSEN, Physics, Yeshiva University
Tuesday, 10 December, 1974

Explanations of Atomicity Before and After Quantum Mechanics
JOAN BROMBERG, History of Physics, B.U. Center for the Philosophy and History of Science
Commentator: PATRICK A. HEELAN, Philosophy, SUNY, Stoney Brook
Tuesday, 21 January, 1975

Gödel's Theorems and Church's Thesis: A Prologue to Mechanism
Judson Webb, Philosophy, Boston University
Commentator: Hilary Putnam, Philosophy, Harvard University
Tuesday, 28 January, 1975

On The Historical Development of the Interpretations of Quantam Mechanics
Max Jammer, History of Physics, Bar-Ilan University (Ramat-Gan, Israel) and CUNY, Queens College
Commentators: Arthur Fine, Philosophy, University of Illinois, Chicago Circle.
Eugene Wigner, Physics, Princeton University
Tuesday, 11 February, 1975

Methodological Reflections on the Theory of Socialist Society
Jerzy Wiatr, Sociology, Institute of Philosophy and Sociology, Polish Academy of Science (Warsaw)
Commentator: Erazim Kohak, Philosophy, Boston University
Tuesday, 18 February, 1975

The Concept of Revolution and the Newtonian Natural Philosophy
I. Bernard Cohen, History of Science, Harvard University
Commentator: Dudley Shapere, Philosophy, University of Illinios, Urbana
Tuesday, 25 February, 1975

Teleology in Machine and Beast: A Reductionist Account of Purposive Behaviour
Roger J. Faber, Physics, Lake Forest College
Commentator: William Pohle, Philosophy, CUNY, Lehman College
Tuesday, 11 March, 1975

Historical Narrative: The Rhetorical Structure of Hegel's Logic
Hayden White, History, Center for the Humanities, Wesleyan University
Commentator: Andrew Arato, Philosophy, Cooper Union
Tuesday, 18 March, 1975

Philosophy of Science 2001
Paul Feyerabend, Philosophy, University of California, Berkeley
Tuesday, 25 March, 1975

Intentions of Intentionality
Jaakko Hintikka, Philosophy, Academy of Finland (Helsinki) and Stanford University
Commentator: Ruth Marcus, Philosophy Yale University
Tuesday, 1 April 1975

Objectivity in Social Science
Paul Diesing, Political Science, SUNY, Buffalo
Commentator: Alasdair MacIntyre, Philosophy and Political Science, Boston University
Tuesday, 8 April, 1975

Truth, Fallibility and The Growth of Knowledge
Isaac Levi, Philosophy, Columbia University
Commentators: Avishai Margolit, Philosophy, Hebrew University of Jerusalem and Harvard University and Israel Scheffler, Philosophy, Harvard University
Tuesday, 22 April, 1975

Philosophical Interpretations of the Principle of Relativity
M. A. TONNELAT, Theoretical Physics, Institut Henri Poincaré (Paris)
Commentator: GERALD HOLTON, Physics, Harvard University
Tuesday, 29 April, 1975
Historical Epistemology and Scientific Practice
MARX W. WARTOFSKY, Philosophy, Boston University
Commentator: ROBERT S. COHEN, Physics and Philosophy, Boston University
Tuesday, 6 May, 1975

Regular meetings (except March 25) to be held at 7.45 p.m. in room 314 of the Sherman Union, 755 Commonwealth Ave., at Boston University Bridge.
For further information consult Professor Robert S. Cohen, Department of Physics, Boston University.

CORRECTIONS

The Editors would like to draw attention to the following misspellings which occurred in the last two issues of this *Journal.*

In A. W. Coats's discussion note 'Situational Determinism in Economics: The Implications of Latsis's Argument for the Historian of Economics' (Vol. 25, No. 3) Dr. Latsis's name is consistently misspelt 'Lastis'.

H. E. Kyburg's name was misspelt 'Kyberg' at the foot of his review article 'Propensities and Probabilities' (Vol. 25, No. 4) as well as in the Contents and in the running pageheads.

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER

MARCH 1975, VOLUME XXXV, No. 3

CONTENTS

I. ARTICLES

Alienation: A Conceptual Analysis, by TRONN OVEREND
Essentialism and Possible Worlds, by JAMES E. TOMBERLIN
On the Relation of Natural Properties to Normative and Evaluative Properties, by PHILLIP T. MONTAGUE
Direct Perception, by D. D. TODD
Heidegger's Theory of Metaphysics and Dialectics, by JAMES LAWLER
Hume, Price, and Testimony, by R. A. NAULTY and P. J. SHEEHAN

II. DISCUSSIONS, REVIEWS
III. COMMUNICATION
IV. NOTES AND NEWS, RECENT PUBLICATIONS

Subscription prices $8.00 per year for individuals; $10.00 per year for institutions. Separate issues, $2,00 and $2.50 respectively. Postage and handling for back issues only. Papers submitted for publication will not be returned unless accompanied by a self-addressed, stamped envelope, or equivalent postage.

A Circular listing the main contents of the Journal since 1940 will be sent upon request.

Address all communications to

PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH

State University of New York at Buffalo,
Buffalo, New York 14226

THE
REVIEW OF METAPHYSICS

A PHILOSOPHICAL QUARTERLY

Edited by JUDE P. DOUGHERTY

Vol. XXVIII, No. 3 March 1975 Issue No. 111

ARTICLES	
Hans-Georg Gadamer	*Inverted World;* translated by John F. Donovan, S.S.
Hugh J. McCann	*Trying, Paralysis, and Volition*
Nicholas Rescher and Thomas C. Vinci	*The Truth-Relevancy of the Pragmatic Utility of Beliefs*
Rudiger Bubner	*Kant, Transcendental Argument and the Problem of Deduction*
DISCUSSION	
Alvin Plantinga	*On Mereological Essentialism*
Roderick M. Chisholm	*Mereological Essentialism: Some Further Considerations*
CRITICAL STUDIES	
Victor Gourevitch	*Rawls on Justice*
James A. Ogilvy	*Review of recent Hegel Studies*
BOOKS RECEIVED	
William A. Frank and Staff	*Summaries and Comments*
PHILOSOPHICAL ABSTRACTS	
ANNOUNCEMENTS	

Individual Subscription $9.00 Student Subscription $5.00
Institutional Subscription $15.00

THE CATHOLIC UNIVERSITY OF AMERICA, WASHINGTON, D.C. 20064
Twenty-Year Cumulative Index, 1947-1967 is available for $3.00 a copy

INQUIRY

An Interdisciplinary Journal of
Philosophy and the Social Sciences
Edited by Arne Naess
and Alastair Hannay

Number 1 *Spring 1975* *Volume 18*

Adaptivity and Self-Knowledge	Amelie Oksenberg Rorty
Moral Black- and Whitemail	H. J. N. Horsburgh
Marx's Commodity Fetishism	Terrell Carver
Grammatical Rules and Explanations of Behavior	Robert E. Sanders and Larry W. Martin
The Range of Intentions	Donald Gustafson
Discussions:	
Agassi on Rationality and the *Tu Quoque* Argument	Robert E. Innis
Nietzschean Notes on the Tennessen-Naess Exchange	[Stanley Paluch
Review Discussion:	
Gene Sharp's *The Politics of Nonviolent Action*	H. J. N. Horsburgh

Published quarterly by Universitetsforlaget, P.O. Box 307
Blindern, Oslo 3, Norway

Annual Subscription: Institutions – N. kr. 75.00, £4.90, $15.00
Individuals – N. kr. 60.00, £3.90, $12.00
Postage included

MIND

A Quarterly Review of Philosophy

edited by Professor **DAVID W. HAMLYN**

Vol. LXXXIV APRIL 1975 No. 334

Articles	
The Similarities Between Space and Time	G. SCHLESINGER
Another Shot at the Canons of Induction	K. S. FRIEDMAN
Expression	D. R. FINN
The Veil of Perception	G. A. J. ROGERS
Analyticity, Necessity and Apriority	R. G. SWINBURNE
Provided Nothing External Interferes	J. MOLINE
Discussions	
Interaction and Physiology	J. BRICKE
Mental Images and Pictorial Properties	S. CANDLISH
Redmon on 'Exists'	S. E. BOËR
A Reply to Boër	R. B. REDMON
Predicating Truth	C. J. F. WILLIAMS
Rationality of Action and Theory-Testing in Popper	A. O'HEAR
Compatibilism and Natural Necessity	N. M. L. NATHAN
A Paradox of Knowing Whether	J. A. BARKER
Scriven and MacKay on Unpredictability and Free Choice	L. D. ROBERTS
Book Reviews	
Books Received	

Annual Subscription £2.50, *Single Issues* 80p

Published for MIND ASSOCIATION by
BASIL BLACKWELL 49 BROAD STREET OXFORD

dialectica

International Quarterly Review of philosophy of Knowledge

Board of Directors (Switzerland)

P. Bernays, Bodmerstrasse 11, 8002 Zürich
Ed. Bertholet, case postale 68, 1880 Bex
F. Bonsack, 155, av. de Cour, 1007 Lausanne
F. Gonseth, chemin du Muveran 12, 1012 Lausanne
D. Jakubec-Vodoz, Av. de Chailly 28C, 1012 Lausanne
H. Lauener, Längfeldstrasse 8, 3063 Ittigen/Berne
B. Morel, rue Ecole-de-Médecine 16, 1205 Genève
P. E. Pilet, avenue Reymondin 13, 1009 Pully-Lausanne
W. Soerensen, "Les Chesaulx", 2036 Cormondrèche/Neuchâtel

Vol. 27/3–4 CONTENTS

Henri Lauener	Zur philosophischen Problematik des Begriffes der Kausalität
E. Kaeser	Ein Dogma des kritischen Rationalismus
Vasile Tonoiu	La philosophie ouverte et le problème du commencement
Paul Gochet	Le problème du commencement et la philosophie analytique
Dennis A. Rohatyn	Against the Logicians: Some Informed Polemics
Charles H. Lambros	Schlicks Doctrine of The A Priori in 'Allgemeine Erkenntnislehre'
Reviewing Studies	

Vol. 28/1–2 CONTENTS

Henri Lauener	Vorwort
Bruno Brülisauer	Wieweit werden philosophische Probleme durch Sprachanalyse gelöst?
Ferdinand Gonseth	Über die Sprache sprechen
Avrum Stroll	Linguistic Clusters and the Problem of Universals
Max Black	Paradigm Cases and Evaluative Words
Erik Stenius	The Problem of AKRASIA as a Subject of Philosophical Analysis
Dagfinn Foellesdal	Indeterminacy of Translation and Under-Determination of the Theory of Nature
Y. Bar-Hillel	Primary Truth Bearers
Donald Davidson	Radical Interpretation
Jaakko Hintikka	Quantifiers vs. Quantification Theory
Paul Bernays	Ueber die Auffassung formaler Systeme als Sprachen
Urs Egli	Semantische Repräsentation der Frage

Distribution: **Dialectica, Case postale 1081, 2501 Bienne (Switzerland)**
Stechert-Hafner Inc., 31 East 10th Street, New York 3, N.Y.
B. H. Blackwell Ltd., Broad Street, Oxford, England

Price: Single Issue $6.00 4 Issues (1 year) $19.00

Zeitschrift für allgemeine Wissenschaftstheorie

Journal for General Philosophy of Science

Edited by ALWIN DIEMER . LUTZ GELDSETZER . GERT KÖNIG

Per year 2 issues of 208 pp. each Annual subscription DM 80,–

From the contents of vol. VI 2

AUFSÄTZE

- John Worrall, Nachruf auf Imre Lakatos
  Imre Lakatos (1922–1974): Philosopher of Mathematics and Philosopher of Science
- Wolfgang Büchel, Die Struktur wissenschaftlicher Revolutionen und das Uhren—"Paradoxon"
- Wolfgang Detel, Zwei Fallstudien zur Prüfung des Falsifikationismus
- Lutz Geldsetzer, Metaphilosophie als Metaphysik
- Bernd Giesen—Michael Schmid, Rationalität und Erkenntnisfortschritt
- Kurt Hübner, Zur Frage des Relativismus und des Fortschritts in den Wissenschaften
- Wolfgang Krah, Zum Falsifikationsprinzip in der Wissenschaftstheorie

DISKUSSION

- Theodor Ebert, Über eine vermeintliche Entdeckung in der Wissenschaftstheorie
- Victor Kraft, Die Giltigkeit von Normen
- Rainer Porstmann, Werturteile als wissenschaftliche Aussagen. Zu Victor Krafts Beiträgen über die Giltigkeit von Aussagen und Normen
- Friedrich Rapp, Kybernetik und Erkenntnistheorie—Bemerkungen zur Konzeption von Georg Klaus
- Robert E. Butts, Philosophy of Science in Canada
- Lutz Geldsetzer, Bericht über den 14. Internationalen Kongreß für Wissenschaftsgeschichte 19.—27.8.1974 in Tokyo.
- Colin Howson—John Worrall, The Contemporary State of Philosophy of Science in Britain

REZENSIONEN

BIBLIOGRAPHIE

FRANZ STEINER VERLAG GMBH . WIESBADEN, GERMANY

Noûs

Nihil philosophici a nobis alienum putamus

A Quarterly Journal of Philosophy

edited by Hector-Neri Castañeda

Philosophy Department, Indiana University

Bloomington, Indiana 47401

NOÛS publishes quality essays and brief discussions on philosophical problems, regardless of the author's philosophical school or point of view. From time to time, NOÛS will publish symposia, surveys of work recently done on specially selected philosophical topics, critical studies of recent books (rather than reviews), and catalogues of recent publications in philosophy. NOÛS publishes the symposium papers of the A.P.A. Western Division meetings in its March issue.

A quarterly, NOÛS is published in March, May, September, and November each year. For *individuals* the subscription rates are: one year, $9.00; two years, $17.00; three years, $25.00. For *bona fide students:* $8.00 per year. For *institutions:* $16.00 per year (plus 10% for subscriptions handled through a subscription agency). Add 50 cents per year for all non-U.S. subscriptions. Make checks for one year subscriptions payable to INDIANA UNIVERSITY-NOÛS. Checks for multiple year subscriptions should be made payable to THE INDIANA UNIVERSITY FOUNDATION-NOÛS.

VOLUME IX NUMBER 1 MARCH 1975

Symposium papers to be read at the meeting of the Western Division of the American Philosophical Association in Chicago, Illinois, April 24–26, 1975.

MICHAEL SCRIVEN—*Causation as Explanation*

ARTHUR FINE—*How to Compare Theories: Reference and Change*

NINO B. COCCHIARELLA—*Second-Order Theories of Predication: Old and New Foundations*

PHILLIP D. CUMMINS—*Berkeley's Ideas of Sense*

FREDERICK A. OLAFSON—*Husserl's Theory of Intentionality in Contemporary Perspective*

PAUL RICOEUR—*Phenomenology and Hermeneutics*

1974 *Referees*

Announcements

Recent Publications in Philosophy

Order from

The Secretary, NOÛS

Department of Philosophy

Indiana University

Bloomington, Indiana 47401

The Monist *An International Quarterly Journal of General Philosophical Inquiry*

Founded 1888 by Edward C. Hegeler *Editor*, Eugene Freeman

Editorial Board: William P. Alston, Monroe C. Beardsley, Lewis White Beck, William A. Earle, Dagfinn Føllesdal, William Frankena, Maurice Mandelbaum, R. Barcan Marcus, Richard Martin, Mary Mothersill, Joseph Owens, Richard Rorty, J. B. Schneewind, Wilfrid Sellars, John E. Smith, Richard Wasserstrom.

Managing Editor, Ann Freeman

EACH ISSUE is limited to articles on a single general topic.
Communicate with the Editor in advance for Special Instructions. Papers should be 6000-8000 words in length and must be submitted in duplicate nine months prior to the scheduled publication of the issue, accompanied by return postage.

GENERAL TOPICS for recent and forthcoming issues:

SCHEDULED PUBLICATION DATES:

Vol. 59,	No. 3	July 1975	Language, Thought, and the World
Vol. 59,	No. 4	Oct. 1975	The Philosophy of Mysticism
Vol. 60,	No. 1	Jan. 1976	Bioethics and Social Responsibility
Vol. 60,	No. 2	Apr. 1976	Philosophy and Religion in the Nineteenth Century
Vol. 60,	No. 3	July 1976	New Directions in Semantics
Vol. 60,	No. 4	Oct. 1976	Historicism and Epistemology
Vol. 61,	No. 1	Jan., 1977	Conceptions of the Self: East and West
Vol. 61,	No. 2	Apr., 1977	The Philosophy of Thomas Reid
Vol. 61,	No. 3	July, 1977	Nominalism, Past and Present

Editorial Office: Box 1908, Los Gatos, California 95030. Business Office: Box 599, La Salle, Illinois 61301

Subscription Rates: United States: Annual (4 issues): Institutions, $10.00; individuals, $7.50. Single copies: $2.75. Foreign Postage: Add 25 cents to single copy rate or $1.00 to subscription rate.

7 NOV 1975

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1974-5

President: Professor J. W. N. Watkins
President-Elect: Professor D. J. O'Connor
Hon. Secretary: Helena Sheiham
Hon. Treasurer: Dr D. A. Gillies
Members: Dr M. Machover, Mr H. Martins, Mr N. Maxwell, Dr E. Zahar

EDITORIAL BOARD OF THE JOURNAL

Professor R. B. Braithwaite
Professor H. Dingle
Professor Sir John Eccles
Professor Sir Harold Jeffreys
Professor W. C. Kneale
Professor S. Körner
Professor M. Polanyi
Professor Sir Karl Popper
Professor G. J. Whitrow
Professor J. H. Woodger

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules :

The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all persons who have been proposed by two members, approved by the Committee, and elected at an ordinary meeting. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £1·00 ($2.50). Those who wish to join should write to the Hon. Secretary, Helena Sheiham, Department of Philosophy, London School of Economics, Houghton Street, London WC2A 2AE.

Brit. J. Phil. Sci. Vol. 26 No. 1 March 1975

The British Journal for the Philosophy of Science

Contents

Articles

1–18 PAUL FEYERABEND Imre Lakatos
19–26 HEINZ POST The Problem of Atomism
27–39 ROBERT C. SOLOMON Doubts about the Correlation Thesis

Discussions

41–47 LARRY DEWITT Consciousness, Mind and Self: the Implications of the Split-Brain Studies
47–49 I. W. ROXBURGH The Intransitivity of Non Standard Synchronisms
49–51 MENDEL SACHS On the Mach Principle and General Relativity
51–60 L. D. GASMAN Myths and X-rays

Review Articles

61–71 JON DORLING The Structure of Scientific Inference (review of M. B. Hesse: *The Structure of Scientific Inference*)

Reviews

73–87 JOHN F. POST on J. L. Mackie: *Truth, Probability and the Paradoxes;* PETER MACHAMER and STEPHEN LUNSFORD on W. R. Shea: *Galileo's Intellectual Revolution;* I. C. MCFETRIDGE on B. Russell: *Essays in Analysis;* JOHN FOX on D. C. Stove: *Probability and Hume's Inductive Scepticism*
88–90 *Programme of Meetings*

7 NOV 1975

Printed in Great Britain at the University Press, Aberdeen

Volume 26 Number 2 June 1975
BJPIA5 26 (No. 2) 91-186 (1975) 0007-0882

The British Journal for the Philosophy of Science

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editors: JOHN WATKINS and JOHN WORRALL

Assistant Editors: COLIN HOWSON and ELIE ZAHAR

Subscriptions
The *Journal* is published quarterly. The subscription price is £7.00 net (U.S.A. and Canada U.S. $22.00) per volume post free; single issues £2.50 net (U.S.A. and Canada U.S. $8.00) plus postage. The subscription is £4.50 (U.S.A. and Canada U.S. $14.00) for members of the British Society for the Philosophy of Science, who should order direct from the publisher. (Membership of the Society is international and is by no means confined to people resident in the United Kingdom.) Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

Notes to Contributors
Articles, correspondence and books for review should be sent to John Worrall at the Department of Philosophy, London School of Economics, Houghton Street, Aldwych, London WC2A 2AE. Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Footnotes, which should be as few as possible, should be typed separately, double-spaced at the end of the text. They should be referred to by consecutive numbering throughout the text, in which they will be inserted in page proof. Bibliographical references should be listed alphabetically at the end. The style of references should follow current *Journal* practice.

Long quotations are printed in small type without quotes, and should be so marked in the typescript. Single quotes should be used to form names of words and for shorter quotations; double quotes for other purposes. Diagrams, tables, and illustrations should be on separate sheets, with their desired position in the text clearly indicated. Symbolic formulae appearing as separate lines should be clearly distinguished in the text, and foreign letters (Greek, German, etc.) indicated in the margin. Material to be printed in italics, especially English letters used as symbols ('*A*', '*x*', etc.), should be underlined. On matters of style and layout, authors should consult CHAUNDY, T. W. *et al.* [1964]: *The Printing of Mathematics*, London: Oxford University Press.

Offprints of articles may be ordered when corrected galley proofs are returned. Fifty are supplied free; the price of extra copies is given on the order form.

© British Society for the Philosophy of Science 1975

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editors.

Back Volumes
Volumes 1-16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17-25 may be obtained from Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Prices on application.

Brit. J. Phil. Sci. 26 (1975), 91–121 Printed in Great Britain

Metaphysics and the Advancement of Science*

by J. W. N. WATKINS

1 *Introduction.*
2 *'Extra-scientific' Metaphysics and 'External' Influences.*
3 *How to Find the Metaphysical Content of a Scientific Theory.*
4 *Quasi-Ramseyfication: the Case of Newton.*
5 *Hard Cores and* M-*Components.*
6 *The Empirical Side of a Major Scientific Advance.*
7 *The Theoretical Side of a Major Scientific Advance.*
8 *Scientific Reductions.*
9 *Against Presuppositionalism.*

1 INTRODUCTION

The counter-revolution against the logical empiricist philosophy of science seems to have triumphed: I have the impression that it is now almost as widely agreed that metaphysical ideas are important in science as it is that mathematics is.

As one who regards himself, not as a colonel, but as a lieutenant in this counter-revolution,[1] I view this success with some satisfaction. One bad result of logical empiricism was that it created a gulf between the philosophy and the history of science. It was impossible, within the restricted language of logical empiricism, to make cognitive sense of the findings of such historians of science as Meyerson, Burtt and Koyré concerning the decisive role that certain metaphysical speculations have played in the advancement of science.[2] There is now a most fruitful *rapprochement* between methodology and history of science.[3]

---

* A revised and expanded version of a Presidential Address to the British Society for the Philosophy of Science, 11 March 1974. Earlier versions were searchingly criticised by various friends and colleagues. I would particularly like to thank Peter Clark, Colin Howson, Noretta Koertge, David Miller, Alan Musgrave, Sir Karl Popper, Peter Urbach, John Worrall, and Elie Zahar. In 4 I am indebted to Ernan McMullin.

[1] On the identity of the leading colonel, see the section entitled 'Who Killed Logical Positivism?' in Popper [1974], pp. 69–71.

[2] The work of these three men seems to have been generally ignored by logical empiricists. Philipp Frank ([1941], pp. 220–1) is a partial exception. Russell in his [1926] took some (rather worried) notice of Burtt's [1924].

[3] I am thinking here primarily of historical case-studies inspired by Lakatos's methodology of scientific research programmes; for example, Zahar [1973], Lakatos and Zahar [1975]. An anthology of such case-studies, edited by Colin Howson, is expected to be published by the Cambridge University Press in 1976.

But there is a danger in the success of this counter-revolution. We should remember that a view of science that becomes fashionable tends to get taken up and misused in outlying areas.[1] For instance, the old idea that the essence of the natural sciences is their precise, mathematical character has, in the past, inspired a good deal of pseudo-scientific work in the social sciences.[2] This might have been checked if there had then been available a good methodological account of the proper role of precision in science. Popperian methodology now provides this: precision is wanted, not for its own sake, but for the sake of increased testability; and since increased precision *may* lead to decreased testability,[3] we have here a way of distinguishing good precision from bad precision.

According to the counter-revolution against logical empiricism, metaphysical ideas can and do play an important, even a central, role in the advancement of science. But this idea is all too liable to misinterpretation and misuse, especially in those quarters where metaphysics tends to be equated with ideology.[4] The likelihood of such misuse might be reduced by a methodological account of the proper role of metaphysical ideas in science. That is what I try to provide in this address. It will turn out that their role is very important and also very difficult to fulfil. The ban on armchair speculation has been lifted, but the occupation of such a chair is very arduous.

---

[1] This happened to Kuhn's [1962]. Miss Masterman, in her [1970], has explained why: before Kuhn, people working in underdeveloped sciences ('fresh contemporary new sciences', p. 69) were worried by their lack of *theory*; after Kuhn, they saw that having no theory does not matter provided you have a paradigm: 'the paradigm is something which can function when the theory is not there' (p. 66).

In the same volume Feyerabend writes: 'More than one social scientist has pointed out to me that now at last he had learned how to turn his field into a "science" . . . The recipe . . . is to restrict criticism . . . and to create a normal science' with one paradigm ([1970], p. 198).

Dr Rose Dugdale, in her 'Economic Theory in Class Society', also draws comfort from Kuhn's views: 'Kuhn has cast serious doubts upon this paragon of the virtues of objectivity—the natural sciences. After all it is not at all clear that science advances as Popper would have us believe' (in Pateman (*ed.*) [1972], p. 166). I owe this quotation to T. W. Hutchison.

[2] Robbins in his [1935] commented on a work on business cycles: 'after a prolonged and valuable account of the course of business fluctuations in different countries since the end of the eighteenth century, an average is struck of the duration of all cycles and a logarithmic normal curve is fitted by Davies's Method to the frequency distribution of the 166 observations involved. What possible meaning can inhere in such an operation?' (p. 113).

[3] A testable hypothesis of the form $\forall x\ (Fx \rightarrow Gx)$ becomes less testable if the predicate F is replaced by a more precise predicate F' such that $\forall x\ (F'x \rightarrow Fx)$. See Popper's comment ([1934], p. 271) on Keynes [1921].

[4] See, for example, Joan Robinson [1962], chapter 1 and *passim*.

2 'EXTRA-SCIENTIFIC' METAPHYSICS AND 'EXTERNAL' INFLUENCES

Before proceeding to the main task, I have to deal with some preliminary problems. The chief of these is created by a serious deficiency in an earlier treatment of metaphysics in my [1958]. I was concerned there, as I am now, with metaphysics *vis-à-vis* science, and especially with metaphysical ideas that have been influential within science. But I was handicapped by the fact that my way of characterising metaphysics applied only to single statements. (I suggested that many typical metaphysical doctrines can be epitomized by a special kind of 'all-some' statement.) And this meant that I could catch hold of a metaphysical idea only so long as it was outside science. I could not pursue it into science. I could differentiate between, say, 'Every event has a cause' and 'All ravens are black', but I could not differentiate between the metaphysical and the empirical content of a scientific system, such as Newtonian theory. My treatment was lop-sided: I could handle what I called 'extra-scientific' metaphysical speculations but not intra-scientific metaphysics.

The reason was twofold. Suppose, first, that a metaphysical statement m_1 gets incorporated into a scientific theory h. It may well be that h contains several premisses which, though each makes its respective contribution to the testable empirical content of h, are separately untestable. Now if m_1 figures essentially in h and is not just an idle addition, then its elimination from h would lead to consequential eliminations and weakenings which might seriously reduce the testability of h. So I did not see in 1958 by what title I could single out m_1 as a distinctively *metaphysical* component of h. I could regard it only as an ex-metaphysical idea which had now achieved scientific respectability within accepted science.

Or suppose, secondly, that m_1 does not get incorporated into h but that, on the contrary, there is in h an idea m_2 that *conflicts* with m_1. (For instance, if h were Newtonian theory and m_2 were corpuscularianism, then m_1 could be a metaphysical theory, such as Cartesian etherism, which says that matter is continuous.) In 1958 I would have liked, taking the metaphysical m_1 as a starting point, to proceed from it to the identification of its metaphysical counterpart in h. But I did not see how I could do this. If I simply took inconsistency with m_1 as my criterion I would get too much; for if m_1 is inconsistent with a part of h, namely m_2, it is inconsistent with h as a whole. I might have tried looking for the weakest statement in h just strong enough to be inconsistent with m_1;[1] but then I would, typically, have got too little, namely the contradictory of m_1 when I wanted to catch the rival 'all-some' statement m_2. (Using this method I might have caught 'Some bodies are composed of indestructible corpuscles' when I wanted to catch 'All bodies are composed of indestructible corpuscles').

---

[1] I owe this suggestion to Elie Zahar.

My concentration on 'extra-scientific' metaphysics led to an interpretation being put upon my [1958] which startled me.

Lakatos in his [1971] drew an important distinction between *internal* developments and *external* factors in the historiography of science. An internal change is one that the historian can reconstruct as coming about in a rational way. If one methodology involves a theory of scientific rationality which obliges the historian to treat as external something that another methodology would have enabled him to treat as internal then, according to Lakatos, that is a *prima facie* argument for the superiority of the second methodology.

He claimed that his own methodology of scientific research programmes enjoyed just such a superiority over Popper's falsificationist methodology; and part of his claim here concerned their respective treatment of metaphysics. In his account, a scientific research programme involves a *hard core* of fundamental postulates which is reflected in a *heuristic* which indicates how subsidiary assumptions not in the hard core should be revised when the theory, as presently constituted, runs into empirical difficulty. Thus the hard core will persist through a series of progressively modified theories, and it is 'metaphysical' (though Lakatos abandoned this term)[1] at least in the sense that it is rendered unfalsifiable by a policy of imputing empirical failures to subsidiary assumptions.

Given this, and given my lop-sided concentration in [1958] on extra-scientific metaphysics, he was able to claim with some justice to have internalized something which Popperians treat as external:

> What for Popper, Watkins and Agassi is *external*, influential metaphysics, here [in the methodology of research programmes] turns into the *internal* 'hard core' of a programme.[2]

Speaking only for myself, I may say that I had regarded it as a merit of my [1958] that it showed that a metaphysical statement may be logically inconsistent with, and hence may provoke rational criticism of, an accepted scientific theory. By way of example, consider Leibniz's deployment of his principle of continuity against Descartes's rules of motion and against Newton's corpuscularianism. That principle says, in part, that there are no 'leaps' in nature; thus if y is a variable physical magnitude which is functionally dependent on some other variable *x*, then a curve relating values of *x* to values of *y* must be everywhere differentiable: it must not be angular or 'kinky'; there can be no singularities or critical points. As thus stated, and taken in isolation, this principle is metaphysical in the sense of being at once non-analytic and empirically irrefutable: it disallows critical points, but it allows critical *regions* and it puts no limit on the smallness of a critical region, provided it remains finite. Thus a dramatic change in the value of *y* might result from a change in *x* which, though finite, is so tiny as to be empirically indiscernible.[1]

---

[1] Lakatos [1970], p. 184.

[2] Lakatos [1971], p. 99; and see pp. 98, 103, and 124, note 26.

Leibniz showed that Descartes's second and third rules of motion, taken in conjunction with his first, violate this principle.[2] He also showed that this principle is violated by the idea of an absolutely hard atom.[3] Supposing that Leibniz was here criticising parts of the accepted science of his day and that his criticisms were taken up and led to changes within science, I would have said in the terminology of my [1958] that an extra-scientific metaphysical idea had here proved scientifically influential. But I certainly would not have meant by this that those changes were externally induced in the sense that they cannot be rationally reconstructed and are amenable only to socio-psychological analysis. For in this example it was a *proposition*, at once highly plausible but untestable and at logical variance with testable scientific theory, which led in an intelligible way to a critical re-examination and revision of accepted scientific ideas.

In short, if Lakatos's terminology had been available to me then, I would have said that I wanted to give an ***internalist*** account of the influence of metaphysical ideas on scientific developments.

---

[1] Leibniz himself emphasised that there are *apparent* discontinuities in nature: 'Everything goes by degrees in nature, and nothing by leaps . . . But the beauty of nature . . . demands the appearance of leaps . . .' (*New Essays concerning Human Understanding*, trans. A. G. Langley, p. 552).

[2] Descartes's rules concern the motions, after collision, of two bodies, *B* and *C*, approaching each other in opposite directions.

Rule 1 says that if *B* and *C* are equal and their velocities are equal, then they will rebound with equal velocities.

Rule 2 says that if their velocities are equal but *B* is greater than *C*, then *B* will continue in the same direction at a reduced velocity, pushing *C* before it.

Rule 3 says that if they are equal but the velocity of *B* is greater than that of *C*, then *B* will continue in the same direction at a reduced velocity, pushing *C* before it.

In his criticism of rules 2 and 3 Leibniz supposes the excess of *B* over *C*, or of *B*'s velocity over *C*'s velocity, to be 'gradually diminished until it passes over into full equality'—at which point *B* and *C* should be at rest by these two rules, contrary to rule 1. See Leibniz, 'Critical Thoughts on . . . the Principles of Descartes' in L. E. Loemker (*ed.*) *Philosophical Papers and Letters* (Chicago University Press, 1956), pp. 657 f.

[3] 'This [principle that no change takes place by a leap] posited, it follows also that atoms cannot be allowed . . .' See Leibniz, 'Specimen Dynamicum' (in Loemker *op. cit.*, pp. 730 f.) No doubt, Leibniz assumed that Newtonian atoms were supposed to be absolutely hard. But were they? In Query 31, in the 1717 edition of the *Opticks*, Newton declared them to be 'incomparably harder' than any body compounded of them (Dover edition, 1952, p. 400). But he also said that 'Bodies which are . . . absolutely hard . . . will not rebound from one another' (p. 398). Do Newtonian atoms rebound?

But quite apart from its suggestion that I took an externalist view of the influence of metaphysics on science, my ability to handle only 'extra-scientific' metaphysics left a serious gap in my treatment. If there is metaphysics inside science, then there must be extra-scientific metaphysics consisting of untestable propositions conflicting with intra-scientific metaphysics. But the converse does not hold, as we have seen. The existence of extra-scientific metaphysics is consistent with science being metaphysics-free. I try to close this gap in the next section. Faraday confessed himself 'unable to define what is metaphysical in physical science' (see Elkana [1974], p. 169). I will now suggest a way to do this.

3 HOW TO FIND THE METAPHYSICAL CONTENT OF A SCIENTIFIC THEORY

My idea here is really very simple. There are well-known methods, due to Ramsey and Craig,[1] for emptying the metaphysical or theoretical content out of a scientific theory: one replaces the theory by an empirically equivalent surrogate whose predicates are all empirical or observational. My idea is to use one of these methods, namely Ramsey's, in reverse: having Ramseyfied a theory we now ask how this Ramseyfied version of it should be expanded to make it equivalent to the original full-blooded theory, the excess content of the theory over and above its empirical content being regarded as its metaphysical content.

A logical empiricist would protest that what Ramsey's and Craig's methods reveal is the *redundancy* of theoretical terms: the content of a theory just *is* its empirical content.

Later, I will give what seems to me a telling argument against the equation of a theory's content with just its empirical content. In the meanwhile I will only say that to claim that Ramsey's (or Craig's) method has rendered *theories* redundant seems to me like claiming that the invention of high-fidelity stereophonic recording techniques has rendered orchestras redundant. We must first have the full-blooded theory before we can make a Ramsey (or a Craig) recording of its observational content.[2] Something important gets left out when a rich scientific theory is Ramseyfied; and this empiricist technique can be used to *identify* what gets left out.

---

[1] See Craig [1953], [1956]. For Ramsey, see p. 99, footnote 1, *below*.

[2] This point is not new. 'Both Craig's method and the Ramsey device must operate upon theories . . . which are "already there". They eliminate theoretical terms only after these terms have already been used . . . Consequently neither provides a method for the elimination of theoretical terms in the all-important "context of discovery".' (Maxwell [1962], p. 17.) Maxwell's objection is discussed in Cornman [1972], pp. 115 f.

The Ramsey method requires a division of predicates into empirical and theoretical. This division is bound to be somewhat arbitrary, since all 'empirical' predicates have some theoretical or dispositional meaning.[1] And this means, as Feyerabend has so often insisted,[2] that people with differing theoretical outlooks may attach different meanings to a given 'empirical' term. For our purposes, this will not matter in cases where the term, nevertheless, has for them the same extension. In other cases difficulties may be mitigated by observing the following rule. If in a discussion you find that an 'empirical' term is being so differently interpreted that it has different extensions for participants in the discussion, try to find the common area of these different extensions and then agree upon a new term which is intended to have *that* as its extension. If the same trouble recurs with this new term, repeat the process; and so on.[3]

Let us suppose that 'F' and 'G' are taken as empirical predicates with agreed extensions, while 'P' and 'Q' are theoretical predicates. Now consider a simple theory T consisting of these four postulates:

(1) $\exists x\ Px$

(2) $\forall y\ (Py \rightarrow Qy)$

(3) $\forall y\ (Fy \rightarrow Py)$

(4) $\forall y\ (Qy \rightarrow Gy)$

Clearly, T entails the following empirical law-statement:

(5) $\forall y\ (Fy \rightarrow Gy)$

It may be asked why I have included the purely existential postula (1) in T, which is not needed for the derivation of (5). My reason is thi Without (1) T would be non-committal as to the existence of the theoretica entities in terms of which it explains (5). Now I hold that a scientifi theory, taken realistically, does typically affirm the existence of whatever entities it postulates in its explanations of empirical regularities. For instance, molecular theory surely affirms the existence of *molecules* when it explains, say, the condensation of a gas, as the temperature decreases, first into droplets, then into a liquid, then into crystals, and then into a solid —when it explains all this in terms of the decreasing mean kinetic energy of the molecules and the increasing effect of intermolecular forces on them.

---

[1] See Popper [1934], p. 95.

[2] '. . . the meaning of observation sentences is determined by the theories with which they are connected. Theories are meaningful independent of observations; observational statements are not meaningful unless they have been connected with theories.' (Feyerabend [1965], p. 213.)

[3] Feyerabend's incommensurability thesis ('each theory will possess its own experience, and there will be no overlap between these experiences', *ibid.* p. 214) implies that this process would never come to an end, but I see no reason to believe that. Giedymin [1973] has pointed out that two theories (such as Newtonian mechanics and Special Relativity) which have been alleged to be incommensurable because of the conceptual disparity between them may be rendered logically comparable if Ramsey-sentences for each of them can be constructed.

Before turning to a Ramseyfied version of T let us ask what an empiricist might take as the empirical content of T. One answer which he might give is that it is the set of all those logical consequences of T that are non-analytic and *expressible in an empiricist language* (that is, a language whose extra-logical vocabulary contains only empirical predicates).[1] A second, stronger answer that he might give would add that they must also be *empirically testable.*

I will now give the argument referred to earlier against the logical empiricist's identification of factual content with *empirical* content. Carnap agreed with Tarski that the factual content of a sentence is its consequence-class, the class of all its non-analytic logical consequences.[2] But the empirical consequences of T do not constitute a consequence-class on either of the above two answers. One logical consequence of T is:

(6) $\exists x\ Gx$.

This will be empirical by the first answer (it is non-analytical and its sole predicate is empirical). But a consequence of (6) is:

(7) $\exists x\ (Gx \vee Qx)$;

and (7) is not empirical by the first answer (the predicate Q is not empirical).

According to the second answer, the purely existential (6), being untestable, is not empirical. Thus the disjunction of (5) and (6) will be a non-empirical consequence of (5); but (5) is empirical (it is non-analytic, all its predicates are empirical, and it is testable).

Thus according to both answers, T has logical consequences which (i) are empirical and which (ii) have non-analytic consequences which are not empirical—so that the empirical content of T does not constitute a consequence-class. Of course, other definitions of 'empirical content' are possible; but so long as 'empirical' is defined more strictly than 'non-analytic' it seems that this result will continue to hold. A theory may have factual content in excess of its empirical content. Henceforth, I will refer to this excess content as the theory's M-component.

I propose to try to define the M-component of our T more effectively with the help of a Ramsey-sentence for T.[1] I will call this sentence T_R. To construct it we first re-formulate T as one conjunctive statement:

---

[1] See Hempel [1950], section 3.

[2] Carnap [1937], pp. 41–2; *cp.* Tarski [1930].

(T) $\exists x Px \wedge \forall y\,((Py \rightarrow Qy) \wedge (Fy \rightarrow Py) \wedge (Qy \rightarrow Gy))$.

Next, we replace the theoretical predicates 'P' and 'Q' by predicate variables, say 'ϕ' and 'ψ'; finally, we bind these variables by existential quantifiers '$\exists\phi$' and '$\exists\psi$' placed in front of the whole expression. Thus we get:

(T_R) $\exists\phi\exists\psi\,[\exists x\phi x \wedge \forall y\,((\phi y \rightarrow \psi y) \wedge (Fy \rightarrow \phi y) \wedge (\psi y \rightarrow Gy))]$.

We could expand T_R to make it effectively equivalent to the original T by adding, within the square brackets, a clause identifying ϕ with P and ψ with Q. The difference between T_R and a thus expanded version would be analogous to the difference between 'Jane has two properties which inspire John to propose to her' and 'Jane has two properties which inspire John to propose to her, namely wealth and old age'. The trouble with this is that the clause '$\ldots \wedge \phi = P \wedge \psi = Q$' cannot be separated off and asserted independently, just as 'namely wealth and old age' cannot.

I have found that the problem of capturing the excess content of T over T_R in *statements* more difficult than I had anticipated. My first idea was to employ a meta-statement which would spell out the theoretical ontology of T by identifying the predicate variables in T_R with the predicates that they replaced. Such a statement might read:

(T_M^+) If T_R is true, then the statement obtained by dropping its existential quantifiers '$\exists\phi$' and '$\exists\psi$' and replacing the predicate variables 'ϕ' and 'ψ' by, respectively, the predicates 'P' and 'Q' is also true.

---

[1] I take my hat off to the person who first discerned the idea of Ramsey-sentences in the dense thicket of Ramsey's 'Theories' (published posthumously in Ramsey [1931], pp. 212–36; Moore, in a preface, mentioned that this part of the book is specially difficult).

Who was this person? Finding no references to this idea of Ramsey's prior to Braithwaite [1953], pp. 79 f., I conjectured that it was Braithwaite (who, of course, had edited Ramsey's papers); and I asked him if he could confirm this. He replied that although he had got very near to discovering what are now called Ramsey-sentences in Ramsey's [1931], he had not quite done so: he had got the idea of turning a scientific theory into 'one single existential proposition: There exist properties X, Y, Z, *etc.*, which are such that, *etc.*' (p. 81); but he had not explicitly introduced existentially quantified predicate variables. He added that he believed that he had first become acquainted with Ramsey-sentences, as now understood, through the chapter on 'The Ramsey Sentence' in Carnap's [1966].

My feeling that he was being too modest was confirmed when I looked up that chapter. After remarking that few of Ramsey's colleagues gave much attention to this insight of his, Carnap continued: 'One of the exceptions was Braithwaite. . . . In his book, *Scientific Explanation* (1953), Braithwaite discusses Ramsey's insight, emphasizing its importance.' (p. 252.)

There is a lucid critical discussion of Ramsey-sentences in Scheffler [1963], pp. 203 f. See also Hempel [1958] and Cornman [1972].

Actually, the weakest statement which, when conjoined with T_R yields exactly T is:

$$(T_M{}^*) \qquad\qquad T_R \supset T.$$ [1]

But neither T_M^+ nor $T_M{}^*$ fully captures what I regard as the M-component of T. For instance, I would certainly want to include in this the theoretical, existential, and unconditional postulate (1); but this does not follow from either of the conditional statements T_M^+ or $T_M{}^*$.

I came to realise that the M-component of T should be taken as a set of statements which cannot be summed up by a single statement. The reason is this. For every set of statements that does not constitute a consequence-class there is no statement that has just these statements as its consequences. Now we are thinking of metaphysical content as something rather analogous to falsity content: the falsity content of a statement is the set of those of its consequences which remain after the subtraction of all its true consequences, and this set does not constitute a consequence-class.[2] Likewise, if we take the metaphysical content of a theory to be, roughly (I will revert to this in a moment), the set of those of its consequences which remain after the subtraction of all its empirical consequences, this set too will not constitute a consequence-class.

Thus instead of trying to capture the M-component of T in a single statement we shall have instead to determine a *set* of statements that constitutes its M-component. I propose to say that a statement belongs to T_M (the M-component of T) if it is a consequence of T that is neither (i) a consequence of T_R nor (ii) empirically testable. I add requirement (ii) because, without it, T itself would belong to its own M-component. And I do not rely on (ii) alone because, without (i), the M-component of T would include those of its consequences whose predicates are all empirical but which are untestable. Thus we get a tripartite partition of the (non-analytic) consequences of T into: (*a*) those that are metaphysical; (*b*) those that are empirically testable: and (*c*) a residuary class of what, for want of a better name, I will call the quasi-empirical consequences of T. Class (*a*) includes each of the postulates of our T taken singly; it also includes all those conjunctions of them that are untestable. Class (*b*) includes the empirical law-statement (5) above; it also includes

---

[1] Tuomela calls such a sentence a theory's *Carnap sentence* ([1973], pp. 59 and 135 f.; I owe this reference to David Miller). However, Carnap held that such a sentence is 'purely analytic' on the ground that all the factual content of T is in T_R ([1966], p. 270). I hold that it is only the empirical content of T that is in T_R and that the Carnap sentence will be synthetic if T has an M-component.

Over the problem of capturing the M-component I have benefited from intensive discussions with Messrs Howson, Worrall and Zahar.

[2] See Popper [1963], p. 233.

T_R and, indeed, T itself (which is what we want, for T is a testable scientific theory). Class (*c*) includes those consequences of T_R that are untestable, together with all their (non-analytic) consequences. It contains, so to speak, all the degenerate empirical consequences of T. For instance, it includes the purely existential statement (6) above and the still weaker (7).

I will conclude this section by meeting a possible objection to my use of Ramsey-sentences in the identification of M-components.[1] It might be claimed that a Ramsey-sentence reproduces *all* the factual content of the corresponding theory, on the following grounds: the meanings of the theoretical terms of a scientific theory are determined by their logical interrelations within the theory; although a Ramsey-sentence replaces theoretical terms like 'P', 'Q', . . . by existentially quantified variables 'ϕ', 'ψ', . . .; *all the relations* in which 'P', 'Q', . . . stood within the theory are reproduced, though now as relations in which 'ϕ', 'ψ', . . . stand within the theory; since these variables have exactly the same interrelations in the Ramsey-sentence as the predicates had in the original, the predicates do not have any excess meaning over that of the variables; hence the original theory has no excess content over that of its Ramsey-sentence: there will be nothing left if the latter is subtracted from the former.

I agree, of course, that the meanings of the theoretical terms are *constrained* by their interrelations in the theory, but I do not regard them as implicitly *defined* thereby. (Thus I hold that 'force' in Newtonian theory must be understood as denoting something that is measured by mass times acceleration, but I deny that it just *means* 'mass times acceleration'.) The latter view leads to an empiricist or conventionalist interpretation of science: if the primitive terms in a scientific theory are all empirical, all the others being explicitly or implicitly defined, then the theoretical terms are mere convenient abbreviations which could, in principle, be dispensed with. For someone who takes this view, of course, nothing is lost by Ramseyfying a scientific theory.

The thesis that the theoretical terms are *wholly* determined by their positions in a completed scientific theory runs into difficulty when we consider a theory in the process of being constructed. Suppose that some of the theoretical postulates of a future scientific theory have been adumbrated but that these have not yet been connected to empirical propositions. Such a partial sketch, which has as yet no empirical content, may nevertheless constitute a synthetic assertion about the world. But in this case the claim that the meanings of the theoretical terms are determined by their interrelations, and that these are reproduced in the corresponding

---

[1] This objection was put to me independently by Alan Musgrave and Elie Zahar.

Ramsey-sentence, breaks down. A Ramsey-sentence just for postulates (1) and (2) of our T would read:

$$\exists\phi\exists\psi\ [\exists x\phi x \wedge \forall y\ (\phi y \rightarrow \psi y)].$$

And this is, presumably, analytic.[1]

Or consider the question intuitively. Suppose that there had been a theory which explained certain properties of fluids by invoking very small, absolutely hard, spherical, and frictionless atoms. Surely its theoretical terms would have some meaning beyond that imposed on them by their positions in the theory?

I seem to be in disagreement with Popper here. He writes:

> *one* theory T_1 may be formulated in many ways, and may use many different conceptual systems, say C_1 and C_2. Or to put it another way: two theories, T_1 and T_2, should be regarded as one if they are logically equivalent, even though they may use two totally different 'conceptual systems' (C_1 and C_2) or are conceived in totally different 'conceptual frameworks'.[2]

I agree, of course, that if T_1 is indeed logically equivalent to T_2 then they are one and the same theory. But suppose that T_1 and T_2, although they have the same formal structure and yield the same empirical law-statements, have conflicting M-components (as would be the case if rival theoretical interpretations had been placed upon the same formalism). I take it that this, for Popper, would be a case of two theories being 'equivalent even though their "underlying" conceptual frameworks are utterly different'. For me, it would be a case where, although the Ramsey-sentences of two theories are logically equivalent, the theories themselves are *not* equivalent, indeed are logically inconsistent with each other at the metaphysical level.

4 QUASI-RAMSEYFICATION: THE CASE OF NEWTON

Long before Ramsey invented Ramsey-sentences there were scientists who presented their theories, or parts of them, in what I will call a quasi-Ramseyfied form, achieving in an informal way the same sort of ontological agnosticism. A quasi-Ramseyfied theory co-ordinates empirical law-statements by postulating certain unobservable factors but without specifying just *what* those factors are. The theory's M-component is left more or less shadowy. Such quasi-Ramseyfication is intermediate between a full-fledged realist presentation (in which unobservables are postulated *and* specified) and an instrumentalist presentation (in which the theoretical superstructure of the theory is presented merely as so much calculating apparatus to which nothing corresponds in nature).

---

[1] I inserted 'presumably' in deference to those anti-platonists for whom any sentence of the kind '$\exists\phi\exists\psi\ (\ldots\phi\ldots\psi\ldots)$' is *false*. Scheffler ([1963], pp. 210 f.) has examined, with acknowledgments to Bohnert, some ways in which Ramseyfication can trivialize statements that are synthetic but untestable.

[2] Popper [1967], p. 12.

When a scientist comes to publish his theory he may have various reasons for preferring to take an agnostic or cagey line with regard to some or all of its *M*-component.

A good reason would be that he is himself still undecided as to the real nature of the unobservables in question: he may not be ready to offer a thoroughly realist presentation of his theory. Or perhaps he has arrived at a definite view concerning them but judges it to be so counter-intuitive and controversial that he prefers to suppress it, fearing that its avowal might jeopardise the success of his whole theory.

Both kinds of reason seem, at different times, to have encouraged Newton to present a quasi-Ramseyfied version of his theory of universal gravitational attraction.[1]

Newton did not quasi-Ramseyfy the *whole* of the theory of the *Principia*. Had he done so, it might have run something like this.

(1) There are entities *P* such that every body is composed of one or more *P*'s.

(2) Every *P* and every body composed of *P*'s has a certain intrinsic property *I*.

(3) *I* is such that, for any body not moving with uniform rectilinear velocity, there is an external factor *F* causing its deviation from uniform rectilinear motion.

(4) Between any two bodies there is a mutual relation *G*.

(5) For any body in free space not moving with uniform rectilinear velocity the *F*-factor is a function of its *G*-relations with other bodies.

These unspecified things being postulated, the theory would have gone on to say that *P*, *I*, *F*, and *G* are such that . . . and here would have followed a mass of law-statements about the motions of pendulums, projectiles, tides, planets, planetary satellites, comets, *etc.*, which, when conjoined with appropriate statements of initial conditions, would yield an immense array of predictions about spring tides, eclipses, the re-appearance of comets, *etc.*

---

[1] After I had written a first draft of this section on Newton, Ernan McMullin gave an excellent paper, at our *LSE* seminar, on 'Newton's Concept of Matter'. (*I* would like to have re-titled his paper 'An Examination of Newton's attempts to clarify the *M*-component of the *Principia*.') Afterwards, he lent me a typescript. Since I am not a Newton scholar I decided, in re-writing this section, to rely on his scholarship, which included researches into draft material of Newton's in the University Library of Cambridge. McMullin makes special acknowledgments to J. E. McGuire. His paper is due to appear in a collection he is editing: *The Concept of Matter in Modern Thought* (1975).

Actually, in the *Principia* Newton disclosed much of the M-component of his theory. Concerning P he wrote:

> . . . we conclude the least particles of all bodies to be also extended, and hard and impenetrable, and movable and endowed with their proper inertia. And this is the foundation of all philosophy.[1]

The case of I is more complex. As McMullin has shown, at least two ideas of inertia can be detected in the *Principia*, one active, the other passive.[2] According to one, inertia is a *continuously acting* 'force', innate in all bodies, which tends to conserve a body in 'its present state, whether it be of rest, or of moving uniformly forward in a straight line'. According to the other idea, inertia is a 'power of *resisting*' which is innate in every body but which operates only when the body is being acted upon by external forces. (Thus if all bodies except one were annihilated, inertia in the first sense would continue to act, conserving the surviving body in its present state, but inertia in the second sense would be inoperative.) If there is any difficulty in understanding how Newton conceived I it is not due to inexplicitness on his part concerning it.

His account of F is complementary to that of I:

> An impressed force is an action exerted upon a body, in order to change its state, either of rest or of uniform motion in a right line.[3]

Again:

> . . . the deviation of bodies moving in free spaces from rectilinear paths, and continual deflection therefrom towards any place, is a sure indication of the existence of some force which from all quarters impels bodies towards that place.[4]

It is over G that Newton resorted to what I am calling quasi-Ramseyfication and to what McMullin calls deontologising.

> What I call attraction may be performed by impulse or by some other means unknown to me. I use that word here to signify only in general any

---

[1] *Principia*, Book III, Rule III (p. 399 in the Motte-Cajori translation, University of California Press, 1934).

[2] McMullin [1975], section 2.3.

[3] *Principia*, Def. IV (p. 2).

[4] *Principia*, 'The System of the World' [4] (p. 552).

> force by which bodies tend towards one another, whatsoever be the cause.[1]

Later, in connection with the possibility of long-range (and also short-range) forces, he wrote:

> This principle of nature being very remote from the conceptions of philosophers, I forbore to describe it in that book [the *Principia*] lest I should be accounted an extravagant freak and so prejudice my readers against all those things which were the main design of the book.[2]

If McMullin is right, far from resting content with the positivist and mathematical treatment of universal gravitational attraction in the *Principia*, Newton struggled for the rest of his life to sort out this vital part of the *M*-component of his natural philosophy.

5 HARD CORES AND *M*-COMPONENTS

By way of preparation for my view I will first consider some extant views about the role of metaphysics in the advancement of science.

That metaphysics can play such a role was not entirely denied by logical empiricists; but they were debarred by their anti-metaphysical outlook from making any cognitive sense of it: they had to treat the influence of metaphysics on scientific thinking as non-rational, causal, and external. Thus Neurath conceded that metaphysics 'may perhaps act as a stimulus, like some narcotic'.[3] The narcotic-theory of metaphysics is, for me, an unlamented casualty of the counter-revolution against logical empiricism.

When Popper, in his [1933], first opposed his falsifiability-criterion for demarcating science and non-science to the logical empiricists' verifiability-criterion for demarcating sense and nonsense, he claimed that his criterion distinguishes

> with sufficient precision the theoretical systems of the empirical sciences from those of metaphysics . . . without asserting the meaninglessness of metaphysics (which from a historical point of view can be seen to be the source from which the theories of the empirical sciences spring) (p. 314).

My [1958] was directly indebted to this criterion, which allows that there are meaningful statements that are neither analytic nor empirical (for instance, bare negations of universal scientific statements as well as certain 'all-some' statements). Since then, however, I have become sensitive to the charge that, while my view allows that there is good metaphysics outside science (namely, metaphysical speculations that may be destined to *cease being metaphysical* by being incorporated into scientific theories), it does not allow for metaphysics inside science. And it now seems to me that just this was Popper's view in 1933–4. The paper from which I quoted continues: 'In so far as a scientific statement speaks about reality, it must be falsifiable: and in so far as it is not falsifiable, it does not speak about reality.' Taken at face value, this conflicts with my present view. For the *M*-component of a scientific theory, as I have characterised it, is not falsifiable and does speak about reality.

---

[1] *Opticks*, Query 31, quoted by McMullin, *op. cit.*, note 82. McMullin quotes similar passages from pp. 5–6 and 192 of the *Principia*. For quotations from Newton in a similar vein see Koyré [1965], pp. 149 f.

[2] Quoted from draft material by McMullin, *ibid.*, section 2.5.

[3] Neurath [1931–2], p. 300. Since I had been resisting such externalist views in my [1958] I was naturally startled by Lakatos's claim (see 2 *above*) that I too had given an externalist account of the influence of metaphysical ideas on science.

From my present standpoint, Popper's 1933–4 view comes a shade too close to a view with which I certainly disagree, namely that it was only *in its infancy* that science needed metaphysical inspiration. (I like Agassi's crisp retort: 'science is always in its infancy'.[1]) Duhem seems to have held such a view. As a historian of science he stated that all physical theory originally developed from 'almost formless first sketches';[2] which might perhaps be understood as pre-empirical or metaphysical speculations. But he also held that *mature* physical theory is fully 'autonomous and independent of any metaphysical system':[3] science outgrows its metaphysical infancy.

A continuing role, and a role of great importance, was accorded to something like what I am now calling the *M*-component of a scientific theory by what Lakatos called the *hard core* of a research programme (henceforth, *RP*). But the methodology in which he embedded his idea of a hard core contains a thesis which, if correct, would knock out the attempt I shall make to provide a methodological rationale for the role of metaphysical ideas in the advancement of science.[4] My rationale will start out from a consideration of what is required for one scientific theory to be decisively better than another. But according to Lakatos's methodology we cannot appraise theories; we can only appraise series-of-theories (or

---

[1] Agassi [1958], p. 235. That was Agassi's summary of a reply by Maxwell to the claim that 'metaphysical speculation is a thing of the past and that physical science has extirpated it.' Later, Agassi wrote: '. . . I used to resent the hostility toward metaphysics displayed by my physics teachers . . . They derided all metaphysics as the physics of the past; I extol some metaphysics as the physics of the future' ([1964], p. 192).

[2] Duhem [1906], p. 221.

[3] *Ibid.*, p. 219.

[4] Imre Lakatos urged me, early in 1973, to re-examine his methodological papers with a view to pinpointing our disagreements, and I undertook to do so. I think that it was in October that I told him that I had put my finger on the chief disagreement and that I would need to bring it out in my presidential address (due in March 1974). As he was about to leave for America we agreed to postpone discussion of it until I had a typescript for him to read. He died before my typescript was ready.

*RP*s). There would have been no conflict if he had claimed that besides theory-appraisal there is also *RP*-appraisal. But a touch of imperialism appeared in his later writings. At first, he extended Popper's methodology. Later, he attacked it. (His 'Popper$_2$' turned into an anti-Popper.)[1] In Popper's methodology, the emphasis is on corroboration-appraisals of theories. Lakatos came to insist that *RP*-appraisal has to *replace* theory-appraisal.

> The basic unit of appraisal must not be an isolated theory or conjunction of theories but rather a '*research programme*', with a conventionally accepted (and thus by provisional decision 'irrefutable') '*hard core*' and with a '*positive heuristic*' . . .[2]

Previously he had declared it a crucial feature of his methodology

> that it replaces the concept of theory as the basic concept of the logic of discovery by the concept of *series of theories. It is a succession of theories and not one given theory which is appraised* . . .[3]

His arguments against theory-appraisal do not seem cogent to me. Here is one of them:

> But then any scientific theory has to be appraised together with its auxiliary hypotheses, initial conditions, etc., and, especially, together with its predecessors so that we may see by what sort of *change* it was brought about. Then, of course, what we appraise is a *series of theories* rather than isolated *theories*.[4]

That a theory should be appraised together with auxiliary hypotheses *etc.* provides no support for his series-of-theories thesis. Nor, so far as I can see, does his claim (which I do not dispute) that the appraisal of a theory involves comparison with its predecessor(s). Popperian corroboration-appraisals are also comparative: if a new theory is to gain acceptance, it should be at least as well corroborated as its predecessor. But this does not mean that what we appraise is a series-of-theories. If I say that J. S. Mill was a better utilitarian than James Mill, I am not appraising a series-of-Mills; I appraise one Mill relative to another.

Another argument with which Lakatos defended his shift from 'the problem of how to appraise *theories* to the problem of how to appraise *series of theories*'[1] had to do with his replacement of what he called the *monotheoretical* model of empirical falsification (in which a single theory is confronted by a hard piece of recalcitrant evidence) by a pluralistic model.[2] He pointed out, as Duhem, Popper, and others had done before, that since experimental 'counter-evidence' is theory-impregnated, we actually have to choose between competing theories, namely the theory under test and the theories involved in the interpreted counter-evidence; and he added that in any case the theory under test will not be regarded as falsified 'until we have a better one.'[3]

---

[1] Lakatos in his [1968] tended to ascribe his departures from Popper to someone called 'Popper$_2$'. (*E.g.*: 'Since Popper$_2$ always appraises theories by comparing them, the appraisal is rather of a *series of theories* than of an isolated *theory*', p. 164; again: 'But Popper$_2$ can easily get rid of Popper$_1$'s untenable falsificationist, elimination rule . . .', p. 165.) But who was Popper$_2$? In his [1970] 'the real Popper consists of Popper$_1$ together with some elements of Popper$_2$' (p. 181). By the time of [1974*a*] the real Popper is identified 'essentially' with Popper$_1$ (p. 244). It seems that Popper$_2$ is really the Lakatos who 'presents a very different picture of the game of science from Popper's' ([1974*b*], p. 320).

[2] Lakatos [1971], p. 99, his italics.

[3] Lakatos [1970], p. 132, his italics.

[4] *Ibid.*, pp. 117–118.

But pluralism is not serialism; and it is ironic that an argument for *RP*-appraisal as against theory-appraisal should invoke the idea of one *theory* being better than another.

My own view is that theory-appraisal and *RP*-appraisal are not mutually exclusive. I take it that we want our units of appraisal to be as small as possible, so that the process of amendment and revision in science may be as rational as possible. We want to get as far from Quinean holism as we can. (If you are feeling poorly and go to the doctor, you want him to do better than merely tell you that there is something wrong somewhere.) But how small we can actually have our units will depend on the given situation. Some scientific situations are sufficiently clear-cut to allow one theory to be appraised as better corroborated than its rivals or predecessors. Other situations are messy, with all the available theories attended by more or less serious empirical or other difficulties. In such situations it may be impossible to make comparative corroboration-appraisals (all the theories are discorroborated) but possible to appraise the relative progressiveness of the competing *RPs* in which they have been developed.[4] Minus its ban on theory-appraisal, Lakatos's methodology is a valuable extension of the theory-oriented methodological tradition from which he developed it.

How much similarity is there between my present idea of an *M*-component (*MC* for short) and his idea of a hard core (*HC* for short)? There are certain differences. An *HC* must, by definition, persist through a series of theories, but an *MC* need not. Also, the content of an *MC* is objectively determined (once a conventional division of predicates into 'empirical' and 'theoretical' has been agreed), whereas what constitutes an *HC* is determined by scientists' decisions as to which postulates are to be preserved inviolate, predictive failures being blamed on the 'protective belt' of subsidiary assumptions. This presupposes that the *HC* is not separately testable, apart from subsidiary assumptions; and this means that it must fall within the current theory's *MC* (as defined towards the end of 3 above; I am taking it for granted that what I there called the quasi-empirical consequences of a theory would never be included in the *HC*).

---

[1] *Ibid.*, p. 119.

[2] *Ibid.*, p. 129.

[3] *Ibid.*, p. 121.

[4] See for example Urbach [1974]. Urbach is, I believe, the first to apply Lakatos's methodology to the investigation and appraisal of a *contemporary* scientific debate—and with dramatic effect: what had seemed a messy and obscure situation could now be seen clearly with one *RP* out in front, the other trailing ingloriously behind.

In short, the hard core of a research programme should either coincide with, or be contained within, the *M*-component of a theory developed within that research programme.

6 THE EMPIRICAL SIDE OF A MAJOR SCIENTIFIC ADVANCE

Let T now be the best theory, at a certain time, within the most progressive *RP* in a well-developed branch of natural science; and let T' be a subsequent theory that constitutes a ***major scientific advance*** over T. My aim is to discover what we should expect the relation to be between the *M*-components of the two theories. But to do this we must first consider what the relation between their respective empirical contents should be. Let T_E and T'_E be empirical versions of T and T' (obtained, say, by Ramseyfying them). Then how should T'_E stand to T_E?

I shall rely here on Popper's answer, which includes the following desiderata:

(1) T'_E should *go beyond* T_E in the sense that there are questions answered by T'_E that are not answered (or are not answered with equal precision) by T_E, but no questions answered by T_E that are not answered with at least equal precision by T'_E.[1]

(2) T'_E should ***revise*** T_E, which means that T'_E will be strictly inconsistent with T_E.

This second desideratum splits into a less exacting and a more exacting part:

(2.1) T'_E should revise T_E in places where tests have caused difficulties for T_E.

(2.2) T'_E should also revise T_E in places where tests have, so far, caused ***no*** difficulty for T_E.

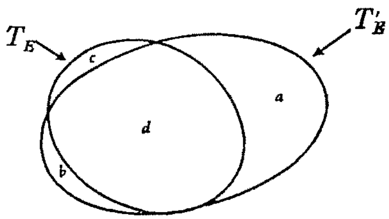
Our (2.2) is one expression of Popper's idea that an explanatory theory should be the reverse of *ad hoc* in relation to its ***explicandum***.[2]

The relation in which T'_E should, ideally, stand to T_E may be represented diagrammatically thus:

---

[1] I added the phrase in parenthesis with the hope of evading an objection by David Miller.

[2] Popper [1972], p. 202.



Here, a is that part of the content of T'_E that goes beyond T_E, b is that part of T'_E that revises T_E, c is that part of T_E that gets revised, and d is their common content.

Of course, if T' is to constitute an actual, and not just a potential, advance over T, T' should go on to win corroborations in the a area, and to be vindicated against T in the b/c areas.

7 THE THEORETICAL SIDE OF A MAJOR SCIENTIFIC ADVANCE

There cannot be any kind of recipe for advancing from the given T to a new T' in a way that complies with the above desiderata. But are there any kinds of consideration that might at least assist in the making of such an advance? Or must T' be invented *ab ovo*?

To prepare the way for my own answer, I will first look at this question from the standpoint of logical empiricism.

According to logical empiricism, any such considerations would have to be logical, or empirical, or methodological; they could not, of course, be *metaphysical.*

For a logical empiricist, desideratum (1) above, taken by itself, might suggest some sort of inductive extension or stretching of T_E; but this seems to be ruled out by desideratum (2), which required T'_E to be logically inconsistent with T_E. It says, in effect, that T_E should be distorted in the process of being stretched.

Desideratum (2.1), by itself, might suggest modifying T_E in a way that takes care of any empirical difficulties that had been besetting it. In other words, (2.1) might suggest a more or less *ad hoc* revision of T_E; but (1) and (2.2) require that compliance with (2.1) should be part of a systematic revision of T_E.

What about desideratum (2.2)? It seems clear that no *empirical* considerations could suggest how T_E should be revised in areas where it has so far been *empirically successful.* Then what about methodological considerations? There is one methodological ideal to which almost all empiricist philosophers subscribe (though they may interpret it differently), namely the ideal of simplicity. Could compliance with (2.2), and indeed the systematic revision of T_E in compliance with all these Popperian desiderata, be inspired by the ideal of greater simplicity?

In answering this it will help if we have before us a standard example of a scientific advance that fits this Popperian pattern. So let T be Kepler's laws and T' be Newtonian theory. Now someone who puts a realist interpretation upon the theoretical content of a scientific theory might reasonably claim that this T' is indeed simpler than this T because it subsumes under a few very general principles a mass of empirical material that is much larger and seemingly more heterogeneous than that subsumed under T.

But we are not now considering this question from a realist standpoint. We are considering it from the standpoint of a logical empiricist who equates T with T_E and T' with T'_E; and if we turn a blind eye to the superstructure of Newtonian theory and attend to its empirical implications for the solar system, we have no reason to claim that these are simpler than Kepler's laws. Given that the planets do not move round the sun with uniform circular motion, what could be simpler and more elegant than Kepler's description of their motion? If we compare this description just for simplicity with the revised Newtonian description (in which the motions of the planets are continuously perturbed and 'the sun is agitated by a continual motion'),[1] we must surely declare the latter to be the less simple. Considered merely aesthetically and without regard to truth, this revised version is a disfiguration of an elegant original.

I conclude that logical empiricism, with its ban on metaphysical considerations, offers no clues as to the sorts of consideration which might help in the advance from T to T'.

Now let us consider our question from the standpoint of the methodology of scientific research programmes.

An exceptional feature of Lakatos's methodology is that it describes *concretely* what kinds of considerations can assist scientists to progress from an existing theory T to a better theory T' (*provided*, however, that they continue to work within the same *RP*). He elaborated his idea of the *positive heuristic* of an *RP* in considerable detail, as well as illustrating it with some striking historical examples. The positive heuristic of an *RP* has precisely the role of assisting in the generation of a progressive series of theories wherein each theory has the same *hard core* as its predecessor while its empirical content has to its predecessor's the relation that our T'_E has to T_E.

In this process, progress is achieved by content-increasing modifications among only the subsidiary assumptions.[2] This raises a question: could such a process continue indefinitely or is there, for a given set of hard core postulates, something like an optimal set of subsidiary assumptions? Well, there is surely no *logical* limit to the process of content-increasing modification. Let me revert for a moment to that artificially simple *T* given in 3 *above*. If we take its purely theoretical postulates (1) and (2) as its hard core and its mixed postulates (3) and (4) as its subsidiary assumptions, then we could continue indefinitely to *stretch* its empirical content by adding subsidiary assumptions analogous to (3) (which would relate ever more empirical predicates with the theoretical predicate '*P*'), and also subsidiary assumptions analogous to (4) (which would relate the theoretical predicate '*Q*' to ever more empirical predicates). And we could at the same time continue indefinitely to *distort* its empirical content by means of small modifications to the empirical predicates. (This consideration underwrites Lakatos's claim that it is always logically possible that a degenerating *RP* will stage a comeback.)

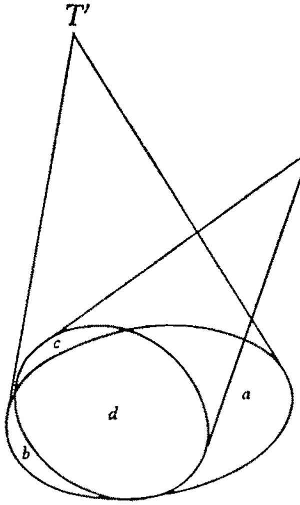
---

[1] *Principia*, Book III, Proposition XII.

[2] Zahar has enriched Lakatos's idea of the heuristic of an *RP* by adding an injunction which might be formulated roughly as follows: if the mathematics of your theory

However, considered historically, it seems that even the best *RP*s eventually run out of steam, that a time comes when the next important advance, if there is one, will come from a theory governed by a different master-idea, a theory which organises the empirical material from a new angle. For Lakatos, this means the emergence of a new *RP* with a different hard core; which means for me the construction of a theory with a radically new *M*-component.

An advance of this kind of *T′* over *T* can be represented diagrammatically by modifying the previous diagram thus:



---

contains entities which appear to be devoid of any physical meaning, try to find a realistic interpretation of them; in this way you may be led to a new physical conjecture. See Zahar [1973], pp. 109 f.

Moreover, there is one kind of major scientific advance, perhaps the most important of all, which can *only* come about in this way. I refer to what are normally called scientific reductions.

8 SCIENTIFIC REDUCTIONS

Let some well-organised department of scientific knowledge have the following character. There is a class of phenomena (it might be heat and temperature phenomena) organised by experimental laws subsumed in turn under a dominant theory T. The phenomena—call them P-phenomena—seem essentially unlike other phenomena, partly because of distinctive empirical properties, partly because of the distinctive interpretation they are given by the ontology of T (which might, for instance, postulate a fluid heat-substance).

I say that a scientific reduction takes place if T is superseded by a new theory T' in such a way that: (1) some or all of the ontology of T' lies outside that of T; (2) the experimental laws derivable from T are derivable (perhaps in slightly revised forms) from T'; (3) phenomena other than P-phenomena are explained by T' in terms of the same ontology. In other words, with the advent of T', the P-phenomena lose their seemingly *sui generis* status: if T' is true, they stand revealed as reducible to something more deep-seated and far-reaching in nature than had been envisaged under T.

A scientific reduction may be radical; or it may be revolutionary. I say that a reduction is radical if T' leaves the ontology of T essentially intact, though showing it to be rooted in a deeper ontology. I say that a reduction is revolutionary if T' replaces much of the ontology of T by a conflicting ontology (replacing, for example, the idea of a fluid heat-substance that behaves deterministically by that of huge numbers of randomly moving molecules displaying very exact statistical regularities).[1]

Revolutionary reductions in science confront the logical empiricist with this dilemma: presumably T can be said to have been reduced to T' only if T can be *derived* from T'; but if T can be logically derived from T', presumably T has not been *reduced* but only subsumed under T'. Nagel has stated this dilemma very clearly. (In the following quotation I replace his expressions 'primary science' and 'secondary science' respectively by my 'T'' and 'T'.)

> . . . a reduction is effected when the experimental laws of [T] (and if it has an adequate theory, its theory as well) are shown to be the logical

---

[1] In my [1974], pp. 395–7, I concentrated on what I now call revolutionary reductions. Peter Clark prompted me to include the less dramatic but still very important case of radical reduction.

> consequences of the theoretical assumptions (inclusive of the co-ordinating definitions) of [*T*] . . . However if the laws of [*T*] contain terms that do not occur in the theoretical assumptions of [*T*] (and this is the type of reduction to which we agreed to confine the discussion), the logical derivation of the former from the latter is *prima facie* impossible.[1]

But the dilemma posed by revolutionary reductions disappears if we distinguish, within the total content of the two theories, a (Ramseyfied) empirical content and a metaphysical content or *M*-component. For we can then say that the requirement of derivability is met in so far as T_E, or a slightly revised version of T_E, is deducible from T', and also that the requirement of reducibility is met in so far as the old ontology of *T* is bypassed in the derivation of (a perhaps slightly revised version of) T_E which is now anchored to the deeper ontology of the more comprehensive T'.

Apart from their tremendous intrinsic interest, scientific reductions are philosophically important in bringing out the ***metaphysical purport*** of great science. This is particularly striking in the case of a revolutionary reduction, which (if the new T' is true) shows that the world, or at least that part of the world which contains the *P*-phenomena, is really very different from what it had appeared to be. So long as the old *T* dominated this domain, these phenomena appeared to fall under a special ontological category. But with the advent of T' the situation undergoes a sea-change. T' undermines the earlier categorisation by relating the *P*-phenomena to an ontology that cuts across old divisions: it reveals these phenomena to be very different, really, from what they appeared to be. Yet T' may disturb T_E very little, or not at all.[2] So T' will also explain *why* they appeared to be other than they really are; it will, to use Popper's term, explain their successful *simulation* of *T*-type phenomena.[3]

Clearly a scientific reduction must involve a significant change in *M*-components. In the case of a radical reduction, T' will postulate an ontology that is *deeper* than that of *T*; in the case of revolutionary reduction, one that is both wider than and *opposed* to that of *T*.

We have now arrived at an answer to the question posed at the outset concerning the proper role of metaphysics in the advancement of science. It is a daunting role which only a rare kind of metaphysical speculation fulfils: to prepare the way for the most important scientific advances of all.

---

[1] Nagel [1961], p. 352. In their [1974] Feigl and Meehl take very much the same view: 'Only if the concepts of the theories (or laws) which are to be reduced (*i.e.* derived, explained) are explicitly definable in terms of the concepts of the reducing theory, can the reduction (derivation) be accomplished' (p. 528).

Nagel tries to resolve the dilemma by introducing into T' assumptions which will satisfy a 'condition of connectability': suitable connections between the terms of *T* and the theoretical terms of T' are to be effected by means of co-ordinating definitions or correlating hypotheses (pp. 353–7). But this would rule out revolutionary reductions.

[2] There *may* be no distorting effect. Popper points out ([1972], pp. 202–3) that Fresnel's wave theory of light was not corrected by Maxwell's electromagnetic wave theory. (John Worrall drew my attention to this.)

[3] Popper [1972], pp. 266 f.

The history of science is usually written mainly in terms of its empirical conquests. I have no complaints about that. But it can be written differently. The first half of Bohm's [1957], which describes the development of modern physics from classical mechanics to quantum mechanics via wave theory, electromagnetic field theory, and kinetic theory, contains few formulas, and says little about experiments. It is a beautifully clear exposition of changing world-views. A history of science that concentrates upon the *M*-components of great scientific theories can unfold a metaphysical drama as exciting and as dialectical as Hegel's (and with the advantage of not having been invented by Hegel).

9 AGAINST PRESUPPOSITIONALISM

In his [1847] Whewell wrote:

> Physical discoverers have differed from barren speculators, not by having *no* metaphysics in their heads, but by having *good* metaphysics while their adversaries had bad: and by binding their metaphysics to their physics, instead of keeping the two asunder.[1]

I like that; but I disagree with Whewell's idea of what constitutes good metaphysics.

There is an approach, deriving from Kant, to the question of good metaphysics which I call ***presuppositionalism.*** Briefly, a 'good' metaphysical principle is, on this approach, one that is presupposed, not just by this or that part of science as it happens to exist at some particular time, but by the very possibility of scientific inquiry. Whewell was such a presuppositionalist. Consider for example his appraisal of the principle of the indestructibility of matter. He allowed that this principle is by no means self-evident to untutored common sense: indeed, daily experience often seems to go against it (as when water evaporates or wood burns). But it is, he held, an indispensable presupposition of scientific thought. Indeed, it is presupposed even in attempts to verify it experimentally.[2]

In a rather similar way, Meyerson regarded a family of conservation principles (inertia, conservation of matter, conservation of energy) as necessary presuppositions or postulates of scientific knowledge: they all derive from the principle of causality; they spell out a proposition whose

[1] Whewell [1847], Vol. 1, p. x.

[2] Whewell [1860], pp. 472 f.

rejection would render scientific inquiry impossible, namely that 'nature is ordered and we can know its course'.[1]

It sometimes happens that the *M*-component of a successful scientific theory contains a part which at first seems so counter-intuitive that it was at first presented in what I called a quasi-Ramseyfied form; and then, as the scientific community becomes reconciled to it, its status is raised until it comes to be regarded as a supreme principle with which all future science must comply. This happened, notoriously, with Newtonian gravitational attraction. Perhaps the man who contributed most to its elevation to what Einstein called 'the Olympus of the *a priori*' was Kant, the greatest presuppositionalist of them all. With his Third Analogy he 'proved' that all substances perceived to coexist in space 'are in thoroughgoing reciprocity'; and in the *Prolegomena* he laid down as a universal principle 'which the understanding knows *a priori*' a 'physical law of reciprocal attraction extending over the whole material nature, the rule of which is that it decreases inversely with the square of the distances . . .'.[2]

By the middle of the nineteenth century the idea of attraction had been elevated into a supreme desideratum for all science. In a famous passage Helmholtz declared:

> Finally, therefore, we discover the problem of physical material science to be to refer natural phenomena back to unchangeable attractive and repulsive forces whose intensity depends wholly upon distance. The solubility of this problem is the condition of the complete comprehensibility of nature.[3]

No one, I take it, would nowadays wish to uphold Kantian presuppositionalism in the face of the tremendous theoretical upheavals in science during the present century. The *empiricial* content of Newtonian theory has been distorted only very slightly by Relativity Theory. But what became of its *M*-component? Euclidean space, absolute time (involving absolute simultaneity), indestructible corpuscles, action-at-a-distance, and (at least in its Laplacean version) mechanistic determinism—all this has been swept aside; and not by the attempts of empiricist philosophers like Mach to purify it of 'idle' metaphysics, but rather by the advent of powerful new theories with new *M*-components.

One could put it like this. According to Kantian presuppositionalism there are certain metaphysical assumptions which *must* remain under the shelter of the umbrella of science so long as science exists; but history teaches us that the umbrella of science lurches about, now sheltering as an

---

[1] See Meyerson [1908], p. 402 and *passim*.

[2] *Prolegomena*, section 38.

[3] Quoted in Einstein and Infeld [1938], p. 58.

M-component an idea which had earlier seemed absurd, now exposing an assumption that had been sheltered for so long that it had come to be regarded as unassailable or even as necessarily true.

However, certain twentieth-century philosophers responded to the overthrow of classical physics by modifying Kantian presuppositionalism into what may be called ***relativistic presuppositionalism.*** Their view runs something like this: True, the presuppositions of science change in the course of time. But if we take a particular scientific epoch we find that certain common presuppositions are being made by the scientific investigators of that time; moreover, relative to that time these presuppositions are *absolute* in that they ultimately and silently control the scientists' thinking and cannot themselves be brought under any critical control.

Collingwood gave such a relativistic twist to presuppositionalism in his [1940]. He saw science as essentially dynamic, as something that must continue to advance or cease to exist;[1] and he recognised that a scientific advance may mean a break with the past ('if quantum theory is true classical physics is false'[2]); so Kantian presuppositionalism is no longer tenable. According to Collingwood's relativistic presuppositionalism, the entire structure of a science during a relatively stable period is determined by the 'absolute presuppositions' of the scientific thinking of that period. These cannot be questioned because they cannot even be avowed by the men who are making them (they can only be formulated afterwards by the historian of science). They undergo change when certain 'strains' become too severe but the process of change is unconscious.[3]

You may say that Collingwood is not nowadays taken very seriously within the philosophy of science. But there are, in Collingwood, anticipations of the more influential ideas of Polanyi and Kuhn. In his [1958], in connection with his idea of the scientist's *commitment* to something *tacit,* Polanyi wrote:

> The curious thing is that we have no clear knowledge of what our presuppositions are . . . When we accept a certain set of pre-suppositions, we may be said to dwell in them . . . Their uncritical acceptance for the time being consists in a process of assimilation . . . They are not asserted and cannot be asserted, for assertion can be made only *within* a framework . . .; as they are themselves our ultimate framework, they are essentially inarticulable.[4]

For Kuhn, that which controls scientific thinking at a given time in a given domain is by no means wholly tacit and inarticulable: a Kuhnian paradigm gets progressively articulated in the course of 'normal science'. But Kuhn's account of the way in which one paradigm gets replaced by another resembles Collingwood's account of the way in which one set of absolute presuppositions gets displaced by another at least in this respect: the process is non-rational (Kuhn likens it to a gestalt-switch).[1]

---

[1] Collingwood [1940], p. 342.

[2] *Ibid.*, p. 259.

[3] *Ibid.*, p. 48, note 1.

[4] Polanyi [1958], pp. 59–60.

From my point of view presuppositionalism, whether Kantian or relativistic, looks like philosophical propaganda for the uncritical retention of part or all of the well-entrenched M-component of prevailing scientific theory. But scientific advances of the kind we have been considering have been made by men who were able, objectively speaking, to burst out of the confines of the seemingly necessary and inescapable ontological assumptions of contemporary science and to develop radical alternatives to them. I say 'objectively speaking' because, subjectively speaking, a revolutionary scientist may remain for some time in the (only slowly relaxing) grip of the old ontological outlook, viewing with considerable suspicion and distrust the alternative ideas that are gradually taking shape in his mind. (Kuhn's idea that the switch 'must occur all at once'[2] seems very wide of the mark.) Someone in such a divided state of mind is likely to appease the more conservative part of his scientific conscience by adopting, at least initially, an empiricist or a quasi-Ramseyfied or an *as if* interpretation of those new ideas. We have seen this in the case of Newton; and it is noteworthy that Einstein, who moved to the extreme right of the positivist-realist spectrum *after* the publication of the General Theory, was still taking an essentially empiricist, Machian view of theoretical science in 1916.[3] And Heisenberg seems to have moved far from the quasi-positivism which had once informed his interpretation of the uncertainty relation.[4] In short, some of those great men who have brought about a radical transformation in theoretical science have tended, at first, to do to their own discoveries something like what Osiander did to Copernicus's, playing down the very ideas which made their achievements possible.

It is not surprising, then, that new departures in science tend to be accompanied by an outbreak of some kind of anti-realist philosophy, whether it be instrumentalism, phenomenalism, or logical empiricism. And it must be admitted that an anti-metaphysical philosophy provides an answer of sorts to presuppositionalism: metaphysical presuppositions of contemporary science cannot be elevated into desiderata for future science if there are no such metaphysical presuppositions anyway.[1] But it provides a wrong answer. Translated into my terminology, presuppositionalism says that if there is an advance beyond an existing theory T to a new theory T', then the M-component of T' must conform with that of T. Logical empiricism says that T has no M-component for T' to conform with. I say that T does have an M-component and that if T' is to effect a revolutionary scientific reduction, then its M-component will conflict with that of T.

---

[1] Kuhn [1962], p. 121. I have discussed Kuhn's view in my [1970].

[2] *Ibid.*, p. 149.

[3] Einstein [1916], pp. 101 f.; quoted by Philipp Frank ([1949], p. 271).

[4] Compare the positivist tone of his comment on the uncertainty relation in Heisenberg [1930], p. 15 and note, with the anti-positivist tone of Heisenberg [1971], chapter 17.

Borrowing Kuhn's distinction between 'normal' and revolutionary science, I would re-write the quotation from Whewell with which this section opened in this way: Physical discoverers have differed from normal scientists, not by having *no* metaphysics in their heads, but by having *new* metaphysical ideas while their adversaries clung to old ones endorsed by existing science; and by binding their metaphysics to revolutionary physical theories instead of keeping the two asunder.

The London School of Economics and Political Science

REFERENCES

Agassi, J. [1958]: 'Koyré on the History of Cosmology', *The British Journal for the Philosophy of Science*, **9**, pp. 234–45.

Agassi, J. [1964]: 'The Nature of Scientific Problems and Their Roots in Metaphysics' in Mario Bunge (*ed.*): *The Critical Approach to Science and Philosophy*. Free Press, Glencoe. Pp. 189–211.

Bohm, D. [1957]: *Causality and Chance in Modern Physics*. Foreword by Louis de Boglie. London: Routledge & Kegan Paul.

Braithwaite, R. B. [1953]: *Scientific Explanation*. Cambridge University Press.

Burtt, E. A. [1924]: *The Metaphysical Foundations of Modern Physical Science*. Second edition 1932. London: Routledge & Kegan Paul.

Carnap, R. [1937]: *The Logical Syntax of Language*. London: Routledge & Kegan Paul.

Carnap, R. [1966]: *Philosophical Foundations of Physics*. Martin Gardner (*ed.*). New York: Basic Books.

Collingwood, R. G. [1940]: *An Essay on Metaphysics*. Oxford: Clarendon Press.

Cornman, J. W. [1972]: 'Craig's Theorem, Ramsey-Sentences, and Scientific Instrumentalism', *Synthese*, **25**, pp. 82–128.

Craig, W. [1953]: 'On Axiomatizability within a System', *Journal of Symbolic Logic*, **18**, pp. 30–2.

Craig, W. [1956]: 'Replacement of Auxiliary Expressions', *The Philosophical Review*, **65**, pp. 38–55.

Duhem, P. [1906]: *The Aim and Structure of Physical Theory*. Translated by P. P. Wiener; Foreword by Prince Louis de Broglie. Princeton University Press, 1954.

Einstein, A. [1916]: 'Ernst Mach', *Physikalische Zeitschrift*, **17**, pp. 101–4.

Einstein, A., and Infeld, L. [1938]: *The Evolution of Physics*. Cambridge University Press.

Elkana, Y. [1974]: *The Discovery of the Conservation of Energy*.

Feigl, H., and Meehl, P. E. [1974]: 'The Determinism-Freedom and Body-Mind Problems', in P. A. Schilpp (*ed.*): *The Philosophy of Karl Popper*. Library of Living Philosophers, Open Court, pp. 520–59.

Feyerabend, P. K. [1965]: 'Problems of Empiricism', in R. G. Colodny (*ed.*): *Beyond the Edge of Certainty*. New York: Prentice-Hall, pp. 145–260.

---

[1] See for example the chapter entitled 'Does Science Have Metaphysical Presuppositions?' in Pap [1949].

FEYERABEND, P. K. [1970]: 'Consolations for the Specialist', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*, pp. 197–230.

FRANK, P. [1941]: *Modern Science and its Philosophy*. Collier Books Edition, 1961.

FRANK, P. [1949]: 'Einstein, Mach and Logical Positivism', in P. A. Schilpp (*ed.*): *Albert Einstein: Philosopher–Scientist*. Library of Living Philosophers, Chapter 9.

GIEDYMIN, J. [1973]: 'Logical Comparability and Conceptual Disparity between Newtonian and Relativistic Mechanics', *The British Journal for the Philosophy of Science*, **24**, pp. 270–6.

HEISENBERG, W. [1930]: *The Physical Principles of the Quantum Theory*. Dover edition.

HEISENBERG, W. [1971]: *Physics and Beyond: Encounters and Conversations*. Translated by A. J. Pomerans. London: Allen & Unwin.

HEMPEL, C. G. [1950]: 'Problems and Changes in the Empiricist Criterion of Meaning', *Revue Internationale de Philosophie*, **11**, pp. 41–63.

HEMPEL, C. G. [1958]: 'The Theoretician's Dilemma: a Study in the Logic of Theory Construction.' Reprinted as Chapter 8 in C. G. Hempel, *Aspects of Scientific Explanation*. Free Press of Glencoe, 1965.

KEYNES, J. M. [1921]: *A Treatise on Probability*. London: Macmillan.

KOYRÉ, A. [1965]: *Newtonian Studies*. London: Chapman and Hall.

KUHN, T. S. [1962]: *The Structure of Scientific Revolutions*. University of Chicago Press.

LAKATOS, I. [1969]: 'Criticism and the Methodology of Scientific Research Programmes,' *Proceedings of the Aristotelian Society*, **69**, pp. 149–86.

LAKATOS, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*, pp. 91–195.

LAKATOS, I. [1971]: 'History of Science and its Rational Reconstructions', in R. C. Buck and R. S. Cohen (*eds.*): *Boston Studies in the Philosophy of Science*, **8**, pp. 91–136.

LAKATOS, I. [1974*a*]: 'Popper on Demarcation and Induction', in P. A. Schilpp (*ed.*): *The Philosophy of Karl Popper*. Library of Living Philosophers, Open Court. Pp. 241–73.

LAKATOS, I. [1974*b*]: 'The Role of Crucial Experiments in Science', *Studies in the History and Philosophy of Science*, **4**, pp. 309–25.

LAKATOS, I., and ZAHAR, E. G. [1975]: 'Why did Copernicus's Programme Supersede Ptolemy's?', in R. S. Westman (*ed.*): *The Copernican Achievement. Forthcoming.*

MASTERMAN, Margaret [1970]: 'The Nature of a Paradigm', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*, pp. 59–89.

MAXWELL, G. [1962]: 'The Ontological Status of Theoretical Entities', in M. Feigl and G. Maxwell (*eds.*): *Minnesota Studies in the Philosophy of Science*, **3**.

MCMULLIN, E. [1975]: 'Newton's Concept of Matter', in E. McMullin (*ed.*): *The Concept of Matter in Modern Thought*. Notre Dame, Indiana: University of Notre Dame Press. *Forthcoming.*

MEYERSON, E. [1908]: *Identity and Reality*. Translated by Kate Loewenberg. London: Allen & Unwin, 1930.

NAGEL, E. [1961]: *The Structure of Science*. London: Routledge & Kegan Paul.

NEURATH, O. [1931–2]: 'Sociology and Physicalism', now in A. J. Ayer (*ed.*): *Logical Positivism*, pp. 282–317.

PAP, A. [1949]: *Elements of Analytic Philosophy*. London: Macmillan.

PATEMAN, T. (*ed.*) [1972]: *Counter Course.*

POLANYI, M. [1958]: *Personal Knowledge: Towards a Post-Critical Philosophy*. London: Routledge & Kegan Paul.

POPPER, K. R. [1933]: 'A Criterion of the Empirical Character of Theoretical Systems', now in Popper [1959], *Appendix* *i, pp. 312–14.

POPPER, K. R. [1934]: *Logik der Forschung*. English translation in Popper [1959].

POPPER, K. R. [1959]: *The Logic of Scientific Discovery*. London: Hutchinson.

POPPER, K. R. [1963]: *Conjectures and Refutations*. London: Routledge & Kegan Paul.

POPPER, K. R. [1967]: 'Quantum Mechanics without "The Observer"', in M. Bunge (*ed.*): *Quantum Theory and Reality*. New York: Springer-Verlag, pp. 7–44.

POPPER, K. R. [1972]: *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press.

POPPER, K. R. [1974]: 'Autobiography', in P. A. Schilpp (*ed.*): *The Philosophy of Karl Popper*. Library of Living Philosophers, Open Court. Pp. 3–181.

RAMSEY, F. P. [1931]: *The Foundations of Mathematics and Other Logical Essays*. R. B. Braithwaite (*ed.*). Preface by G. E. Moore. London: Routledge & Kegan Paul.

ROBBINS, L. [1935]: *An Essay on the Nature and Significance of Economic Science*. Second Edition. London: Macmillan.

ROBINSON, J. [1962]: *Economic Philosophy*. Watts.

RUSSELL, B. [1926]: 'Is Science Superstitious?', reprinted in Bertrand Russell, *Sceptical Essays*, Chapter III.

SCHEFFLER, I. [1963]: *The Anatomy of Inquiry*. New York: Knopf.

TARSKI, A. [1930]: 'Fundamental Concepts of the Methodology of the Deductive Sciences', now chapter V of Alfred Tarski, *Logic, Semantics, Metamathematics*. Translated by J. H. Woodger. Oxford: Clarendon Press, 1956.

TUOMELA, R. [1973]: *Theoretical Concepts*. Wien–New York: Springer-Verlag.

URBACH, P. [1974]: 'Progress and Degeneration in the "I.Q. Debate",' *The British Journal for the Philosophy of Science*, **25**, pp. 99–135 and 235–59.

WATKINS, J. W. N. [1958]: 'Confirmable and Influential Metaphysics', *Mind*, **67**, pp. 344–65.

WATKINS, J. W. N. [1964]: 'Confirmation, the Paradoxes, and Positivism', in M. Bunge (*ed.*): *The Critical Approach to Science and Philosophy*. Free Press of Glencoe. Pp. 92–115.

WATKINS, J. W. N. [1970]: 'Against "Normal Science"', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*, pp. 25–37.

WATKINS, J. W. N. [1974]: 'The Unity of Popper's Thought', in P. A. Schilpp (*ed.*): *The Philosophy of Karl Popper*. Library of Living Philosophers, Open Court. Pp. 371–412.

WHEWELL, W. [1847]: *The Philosophy of the Inductive Sciences*, Second Edition. Parker and Son. 2 Volumes.

WHEWELL, W. [1860]: *On the Philosophy of Discovery*. Parker and Son.

ZAHAR, E. G. [1973]: 'Why did Einstein's Programme Supersede Lorentz's?', *The British Journal for the Philosophy of Science*, **24**, pp. 95–123 and 223–62.

Brit. J. Phil. Sci. 26 (1975), 123–132 *Printed in Great Britain*

Propensity: Popper or Peirce?*

by RICHARD W. MILLER

Recent work by Ian Hacking ([1971*a*] and [1971*b*]) has indicated that there can be value in theories of probability which are no longer in vogue. The justification for such work may lie in the service these theories do in explicating current ones, in the intrinsic merit of the theories themselves, or in both. One such theory, which has seen little detailed analysis, is that of Charles Peirce. For the most part, Peirce's theory of probability has been ignored even by commentators upon his philosophy.[1] Study of Peirce's theory of probability, since it is (as Hacking indicates) dispositional, can shed light upon and perhaps provide a workable alternative to the work of Popper *et al.* on propensity theories of probability.

Evidence of what Hacking calls a 'Peirceian dispositional property' can be found quite explicitly in Peirce's writings. Perhaps the most succinct instance is '. . . the *will be's*, the actually *is's*, and the *have been's* are not the sum of the reals. They only cover actuality. There are besides *would be's* and *can be's* that are real' (8.216).[2] Such a statement bears little similarity to Popper's description of his own propensity theory:

> To sum up, the propensity interpretation may be presented as retaining the view that probabilities are conjectured or estimated statistical frequencies in long (virtual or actual) sequences. Yet by drawing attention to the fact that these

Received 1 *December* 1973

---

* A version of this paper was presented at a meeting of the Western Division—American Philosophical Association on 29 April 1974.

[1] Thayer ([1968]) and Ayer ([1968]) are examples among current writers of those who decide to omit thorough analysis of the theory of probability from a general discussion of Peirce. Burks [1964], Madden [1964], Miller [1970], Tursman [1970], and Will [1965] cast more light upon Peirce's theory (or theories), but problems exist here too: Will and Tursman focus more sharply upon the problem of induction than upon probability itself. Madden is primarily concerned with similarities between Peirce and Reichenbach; my work is an unpublished thesis, thus not easily available; and Burks, it is shown in my [1970], likely is in error in his analysis.

[2] References to works contained in Peirce [1931–58] and in Peirce's unpublished manuscripts among The Charles Sanders Peirce Papers in the Harvard University Houghton Library, will be made in the manner usual to Peirce studies. Thus any reference herein to a decimal number, *e.g.* 5.364, will refer to Peirce [1931–58], volume 5, paragraph 364. References to the unpublished manuscripts will be by either the manuscript or letter folder number assigned them by Professor Richard Robin in his [1967]. Thus *Ms* 339 and L75 will refer to manuscript 339 and Letter folder 75, respectively. Pagination referred to will be as they appear in the *Papers*.

sequences are defined by the manner in which their elements are generated—that is, by the experimental conditions—we can show that we are bound to attribute our conjectured probabilities to these experimental conditions: we are bound to admit that they depend on these conditions, and that they may change with them (Popper [1960], p. 37).

There may, however, be more similarity than first appears. In a 1913 letter to F. A. Woods, Peirce discusses counterfactual conditions, saying:

> . . . if it conveys any meaning, *i.e.* if it is calculated to produce any state of mind, in a person who trusts in it, it must be that it establishes a *habit* in that mind, using the word 'habit' in the original sense, as meaning only that the person or thing has the habit, *would* behave (or usually behave) in a certain way *whenever* a certain occasion should arise. But if this occasion did in actuality *not* arise, such habit of thought as the conditional proposition might produce would be a nullity pragmatistically and practically (8.380).

Thirty-five years earlier (in his Popular Science Monthly series) Peirce had written, 'Belief does not make us act at once, but puts us in such a condition that we shall behave in a certain way, when the occasion arises' (5.373). Of what relevance to probability is a discussion of counterfactuals or a description of belief? Peirce's point, which may in part hinge upon his objective idealism, is that it is through the use of counterfactuals and dispositional, *i.e.* habit-type, statements that we arrive at the concept of the *would be.*

Note also that Popper refers to 'estimated statistical frequences in long (virtual or actual) sequences'. This seems to resemble Peirce's continued reference to the phrase 'in the long run' relative to probability. Peirce's use of the phrase dates as early as his review of John Venn's *Logic of Chance* in 1867 (8.1–6). He does make reference to *will be's* in the Venn review, but a simple grammatical shift of person in his discussion of the solution to a Petersburg-type problem in the same review has him saying: in the long run the average result of either method of repayment *would be* a hundred dollars. It seems a reasonable working hypothesis that this is neither an ambiguity in Peirce's writing nor an equivocal element in his thought, but that it is an attempt to discuss the same idea to which Popper refers as 'long (virtual or actual) sequences'. The *will be* implies the actual sequences; the *would be* the virtual ones. The apparent ambiguity between *will* and *would be* could, therefore, be resolved as an attempt to deal with both actual and virtual sequences.

Most important, however, as an indication of some significant degree of agreement of Peirce with Popper, with regard to conditions being the ultimate source of probabilities, is Peirce's continuing emphasis upon the species/genus relationship upon which probability depends.[1] The genus serves the same function as do Popper's experimental conditions. They both delimit the application of some probability function. Recalling Peirce's association of ***habit*** with ***probability***, we can now clarify how experimental conditions and habits (relative to species/genus distinctions) may be similar. In a letter to Paul Carus, written about 1910, Peirce states:

> If . . . I say that the probability that if a certain die be thrown in the usual way it will turn up a number divisible by 3 (*i.e.*, either 3 or 6) is 1/3, what do I mean? I mean, of course, to state that that die has a certain habit or disposition of behaviour in its present state of wear. It is a *would be* and does not consist in actualities or single events in any multitude finite or infinite. Nevertheless a habit does consist in what *would* happen under certain circumstances if it should remain unchanged throughout an endless series of actual occurrences (L77, 8.225).

The key is, of course, that the habit is 'what *would* happen *under certain circumstances* . . .'. Why not view those certain circumstances, the requisite occurrences for this event to be included within the genus in which we find it, as Peirce's version of Popper's conditions. Then Peirce, like Popper, argues that the source of a consistent probability measure lies in maintaining the genetic (experimental) conditions.

There seem, thus far, to be two major points of agreement about probability between Peirce and Popper: (1) probabilities are physically real relational properties[2] and (2) probabilities are essentially products of consistent experimental setups. These similarities are important but insufficient to show a congruence of the underlying theories.

A difference which, if it cannot be shown to be only apparent, would prove Peirce's and Popper's theories irreconcilable is found in their positions relative to the applicability of probability to single events. Popper's thesis leads him to write:

> The statistical interpretation has been often criticised because of the difficulties of the phrase 'in the long run'. I will *not* discuss this question. Instead I will discuss the question of *the probability of a single event* (Popper [1957], p. 66).

Peirce, in contrast, wrote that '. . . in reference to a single case considered in itself, probability can have no meaning' (2.652).

To see if this apparent difference is, in fact, irreconcilable, it will be necessary to bring it into sharper focus by examining individually the position of each man. Perhaps most critical for the understanding of Popper's position is his paper entitled 'The Propensity Interpretation of Probability'. In it he draws the analogy between propensities and Aristotelian potentialities, while emphasising certain critical differences. Propensities, while physically real, are read as being relational rather than individual, *i.e.* they belong to an experimental arrangement rather than to an individual (Popper [1960], p. 37). Thus, Popper denies that he is offering a new analysis of the term 'probability'; rather, he claims, he is suggesting a new physical hypothesis: '. . . every experimental arrangement (and therefore every state of a system) generates physical propensities which can be tested by frequencies' (Popper [1960], p. 38). In the later paper 'Quantum Mechanics without "The Observer",' Popper further details this position, describing probability as '. . . a *real physical property of the single physical experiment* or, more precisely, of the *experimental conditions* laid down by the rule that defines the conditions for the (virtual) repetition of the experiment' (Popper [1967], p. 33). From this it seems reasonable to infer that he views a probability statement about an event as an elliptical reference to the experimental conditions that gave rise to that event. Since Popper equates those experimental conditions with the single experiment, a probability statement must elliptically refer to a single event. Thus, Popper's belief in the appropriateness of applying probability statements to single (or singular) events (Popper [1960], p. 37) is the consequence of his view of the relationship between probability statements and experimental conditions. The relative frequency statement is simply an empirical means of testing the physical property—the propensity.

---

[1] This emphasis spans, at least, the years 1867–1910. For detail see my [1970], chapter 4.

[2] See Popper [1970], p. 37, lines 7b ff.

In his Harvard University Lectures of 1903, Peirce argued that one has '. . . to specify to what *species* of event it [probability] refers and to what *genus* of event it refers' (5.21). Obviously, if probability belongs to species and genera of events it must be real, and that far Peirce and Popper are in agreement. The closest that Peirce seems to come to an application of probability to singular events is found in a draft of his report to the National Academy of Sciences in 1901. In it he discusses the logic of drawing inferences about history from ancient documents and monuments. Ultimately, he seems to be analysing singular events, *e.g.* in 7.233–255 he applies his techniques to the question of the authenticity of the works of Aristotle. Now clearly we do not, and perhaps cannot, have an actual sequence of worlds in which the works of Artistotle are sometimes authentic and sometimes not, although some works may be authentic and some not in our world. A closer study of his application of probability to such historical events, however, reveals that the probability does not actually attach to the authenticity question itself, rather it applies to the testimony concerning such authenticity. And testimony, given certain parameters concerning the relationship of the witness and that about which he testifies, does fit into the pattern of a long run. Thus, Peirce is seen again to defend the position that 'All that the doctrine of chances can do is to say what will happen in the *long run*' (*Ms* 472, p. 86).

Since Popper's theory derives probability from the experimental setup, his use of the long run is as a test of the probability-hypothesis concerning that set of conditions. What, then, is the function of the long run in Peirce's theory of probability? It must be the device by which a probability-estimate is made and refined, so as to approach the real probability as a limit. The mode of that approach is detailed in his sixth Lowell Lecture of 1903:

> . . . if we take any finite range of conceivable values that includes *p* [*i.e.*] if . . . there will come a time in the succession of the *A*'s after which the proportion of *A*'s that are *Q* among all the *A*'s from the beginning never ceases to be included within that range, then, and only then *p* is the frequency with which the *A*'s are *Q* in the long run, or in other words, is the probability that an *A* will be *Q*. Of course, we cannot say at any time when that time will be or has been after which the value of the ratio in the evergrowing tally will never leave that range of values. For if we could do so, that would constitute a law in the succession and the occurrences would not be independent. But nevertheless the time will come, supposing the occurrences are independent. There must be some ultimate frequency; for otherwise there would be a law of alternation of some kind in the succession (*Ms* 472, pp. 91–93).

The long run, then, has become a kind of limit in itself. Our inability to know when the value of the frequency ratio is set within a range would not preclude its being decided at a quite early stage of tally-keeping. Not having any means, however, of making an evaluation of this probability in the short run, we must look forward to the long run for ultimate verification of what may have been a short-run truth.

We are still left with the question of whether the apparently insurmountable difference in attitude toward the probability of single events can be reconciled. Since Popper is unlikely to be willing to alter his position, its being of the nature of an hypothesis rather than a meaning analysis, Peirce's writings must be sifted for a possible solution. In a draft of a letter to Samuel P. Langley, Peirce writes:

> We can calculate mathematically and therefore deductively precisely how often an induction conforming to certain conditions will give the true probability to a given degree of accuracy, supposing that true probability to be known; and though not precisely, yet within certain limits, how often such an induction will give the probability within certain limits of accuracy even if the true probability is not given. Still, that leaves us entirely in the dark as to the probability that the ascertained probability in any particular case is within any named limits correct. Indeed, it is doubtful whether any meaning can be attached to such a question. All we have to do is to accept the result of induction provisionally, with a degree of confidence depending on the *probable accuracy of the*

> *proceeding*, without troubling our heads about the probability of the inference ratio; and go ahead to get new observations to confirm or modify that ratio (L409, p. 4).

This passage is especially enlightening, for it gives what may be a clue to Peirce's 'plan of operation' for probability in the everyday world. It would seem that the key is 'All we have to do is to accept the result of induction *provisionally* . . .' (*loc. cit.*). If so, then Peirce, the father of Pragmatism, is protected from the accusation of having developed a theory which has no practical consequences and therefore no meaning. If we assume that this is what Peirce intended, then an earlier writing is considerably clarified:

> An individual inference must be either true or false, and can show no effect of probability; and, therefore, in reference to a single case considered in itself, probability can have no meaning. Yet if a man had to chose between drawing a card from a pack containing twenty-five red cards and a black one, or from a pack containing twenty-five black cards and a red one, and if the drawing of a red card were destined to transport him to eternal felicity, and that of a black one to consign him to everlasting woe, it would be folly to deny that he ought to prefer the pack containing the larger proportion of red cards, although, from the nature of the risk, it could not be repeated. It is not easy to reconcile this with our analysis of the conception of chance. But suppose he should choose the red pack, and should draw the wrong card, what consolation would he have? He might say that he had acted in accordance with reason, but that would only show that his reason was absolutely worthless. And if he should choose the right card, how could he regard it as anything but a happy accident? He could not say that if he had drawn from the other pack, he might have drawn the wrong one, because an hypothetical proposition such as, 'if *A*, then *B*,' means nothing with reference to a single case (2.652).

The 'plan of operation' quotation indicates that the reason it would be folly is the *provisional* acceptance of a 25/26 chance of drawing a red card and a 1/26 chance of drawing a black one from the first deck (and the reverse from the second). One then selects an appropriate action *as if* it were an indefinitely repeatable one, making the selection that, assuming the provisionally accepted probability correct, ***in the long run*** would be preferred. In this way we might understand how probability is 'applied' to single and singular events. Thus a man is foolish to ignore what would generally be assumed to be the probability of drawing a red card from the above-described decks, for doing so would be to reject the currently applicable statistics on such events. On the other hand, he would be forced to admit that, the individual being free of other restraints, the outcome of that particular event is unknowable until it has happened, thus making it unreasonable to decry the failure of an event to happen as it 'usually' does.

Is there any way, then, that Peirce could be persuaded to accept Popper's hypothesis about probability applying to single events? I think not. Peirce might formulate this sort of response to Popper's hypothesis:

(1) Although one must be quite careful in the delineation of the generic class in a probability relation, *i.e.* Popper's experimental condition,[1] one ought never to be deluded into thinking that an examination of those conditions alone is sufficient to determine probability. When Popper says, for example, '. . . the probability of 5 turning up will be the same for all dice having only one side marked "5" . . .; and it will be a different one for all dice having more, or less, than one side marked "5", or for non-homogeneous dice' (Popper [1967], p. 19), he is being misled by his use of dice for a model. Despite his professions of reliance upon frequency records,[2] he confidently assigns probabilities based only upon his presumed knowledge of the experimental conditions. While it does appear that one might examine a die, find no evidence to indicate heterogeneity, and thereby hypothesise probabilities of 1/6 for each face, that is not the way one ought to derive a probability. Popper's problem is a simple one. He has examined the experimental conditions and hypothesised that the event has a certain probability. Popper's claim is that we then use the long run statistical analysis to evaluate the hypothesis. Let us assume that the die is, in fact, loaded, but the loading is done in such a way that current technology is inadequate to discover. After a relatively short run, say two or three thousand throws, the loading would surely appear in our statistical record. Popper might, at this point, re-examine the experimental conditions—changing locations to void the possibility of special local conditions, changing operators to void the possibility of *ESP* interference, and so on—but *ex hypothesi* he will find nothing to indicate that the die is not a fair one. Since Popper is limited to a finite number of tests before the die wears enough to affect the experimental conditions, he can never really take his evaluative procedure into account. In contrast, Peirce's provisional acceptance of the statistical record as a basis for action will allow him to utilise what information is available while awaiting further evidence. Pragmatically Peirce's system seems more desirable.

(2) Popper's example which is supposed to put all opposition to final rest is that of a sequence *b* which is composed of throws with a loaded die and of throws with a fair die. The sequence is structured so that members resulting from throws of the loaded die are virtually infinite, while members resulting from throws of the fair die total at most three. The loaded die is assumed to turn up six with a frequency of 1/4. Obviously

---

[1] Popper's view of the relationships between probability and experimental conditions is discussed on pp. 125–6 above.

[2] *E.g.* Popper [1960], p. 38 or [1967], p. 32.

the relative frequency of sixes in b will also be 1/4, there being too few fair tosses to affect the statistical record. Popper claims that our information may be summarised in two equations:

$$(i)\quad p(a, b) = 1/4$$
$$(ii)\quad p(a, bc) = 1/6$$

where a is the set of occurrences of six *on a throw of any die*, b is the set of occurrences which are part of the hypothesised sequence, and c is the set of occurrences of the throws with a fair die. The point which he wishes to prove by this example is that, given any frequency theory which does not apply probability to single events, equation (*ii*) must be false. This is so because bc is very small, which does not allow the appropriate statistical evidence to be gathered. It seems sure that Peirce would have taken Popper to task over the description of the experimental structure. First, there is a fourth class which should have been included among those Popper identified, d, the set of occurrences of throws of a die loaded as the one hypothesised is loaded. This means that we have two equations which seem unproblematic:

$$(i')\quad p(a, d) = 1/4$$
$$(ii')\quad p(a, c) = 1/6.$$

Additionally we have the equations over whose interpretation the problem arises:

$$(iii')\quad p(a, b) = 1/4$$
$$(iv')\quad p(a, bd) = 1/4$$
$$(v')\quad p(a, bc) = 1/6$$

(*iii′*) is true *ex hypothesi*, that is because b is so structured as to be, for all practical purposes, indistinguishable from d. Thus, pragmatically, (*i′*), (*iii′*), and (*iv′*) are the same equation. If we are to accept (*v′*) as true, it must be because we recognise the pragmatic identity of it and (*ii′*), but ***not*** for the same reason we identified the other three. The reason for the identification must lie in our realisation that with regard to (*iv′*) and (*v′*) ***the b is irrelevant***. We differentiate the two types of throws generically as throws of a loaded die and throws of a fair die. There need be no application of probability to the single cases to do that. Thus, when Popper writes:

> . . . if the probability is a property of the generating conditions—of the experimental arrangement—and if it is therefore considered as depending upon these conditions, then the answer given by the frequency theorist implies that the virtual frequency must also depend upon these conditions. But this means that we have to visualise the conditions as endowed with a tendency, or disposition,

> or propensity, to produce sequences whose frequencies are equal to the probabilities; which is precisely what the propensity interpretation asserts (Popper [1960], p. 35).

Peirce would agree wholeheartedly with this statement. Peirce would simply assert that the best that can be expected from Popper's example is support for propensity. The application of probability to single events is neither proven as necessary nor shown to be especially useful. It surely must be shaved by Occam's Razor.

Can Peirce, then, be reconciled with Popper on the question of single event probability? The evidence clearly says, 'No'. As Popper pointed out, the six may never occur on the tosses within such a hypothesised sequence made with the fair die. This is because, Peirce would say, the occurrence or non-occurrence of an event is not taken as either confirmation or disconfirmation of probability, '. . . in reference to a single case considered in itself, probability can have no meaning' (2.652). Popper can, if he wishes, call the provisional assumption of preliminary statistical results a 'new physical hypothesis' but that does not relieve him of the responsibility for justifying a new hypothesis when the old one (*i.e.* Peirce's) was still adequate.

What are we now justified in inferring about propensity as a theory of probability in general, and about the specific versions offered by Peirce and Popper? First, both men seem to give strong support to the proposition that the propensity interpretation is both useful and accurate. This can be seen in Peirce's arguments against *a priori* theories (detailed in Miller [1970]), in Popper's arguments against subjective theories (detailed in Popper [1960]), and in the arguments of each for propensity (detailed above). Second, Popper's schema is defective in that it neither can adequately apply its own accepted test of experience to cases where the statistical results are at variance with the expected results based upon the experimental conditions, nor can it justify the insertion of the hypothesis of single event probability. Third, Peirce's theory can deal with the same questions as does Popper's, but is not liable to the criticisms indicated above. Thus Peirce's theory, although currently out of vogue, is not only a valuable tool for uncovering shortcomings in Popper's theory but seems to be a desirable substitute for it.

University of Missouri–Rolla

REFERENCES

AYER, A. J. [1968]: *The Origins of Pragmatism.* San Francisco: Freeman, Cooper and Co.

BURKS, A. W. [1964]: 'Peirce's Two Theories of Probability', in Edward C. Moore and Richard S. Robin (*eds.*): *Studies in the Philosophy of Charles S. Peirce—Second Series.* University of Massachusetts Press, pp. 141–50.

FISCH, M. H. [1967]: 'Peirce's Progress from Nominalism Toward Realism', *The Monist*, **51**, pp. 159–78.

HACKING, I. [1971*a*]: 'Equipossibility Theories of Probability', *British Journal for the Philosophy of Science*, **22**, pp. 339–55.

HACKING, I. [1971*b*]: 'Jacques Bernoulli's *Art of Conjecturing*', *British Journal for the Philosophy of Science*, **22**, pp. 209–29.

MADDEN, H. [1964]: 'Peirce on Probability', in Moore and Robin (*eds.*): *Studies in the Philosophy of Charles S. Peirce—Second Series*, pp. 122–40.

MILLER, R. W. [1970]: *Charles S. Peirce's Theory of Probability*. (Unpublished doctoral thesis, University of Illinois.)

PEIRCE, C. S. [1931–58]: *The Collected Papers of Charles Sanders Peirce*. A. W. Burks, C. Hartshorne, P. Weiss (*eds.*). Belknapp Press of the Harvard University Press, Cambridge.

PEIRCE, C. S.: The Charles Sanders Peirce Papers. (Unpublished manuscripts in the Houghton Library, Harvard University.)

POPPER, K. R. [1960]: 'The Propensity Theory of Probability', *British Journal for the Philosophy of Science*, **10**, pp. 25–42.

POPPER, K. R. [1957]: 'The Propensity Theory of the Calculus of Probability and the Quantum Theory', in S. Körner (*ed.*): *Observation and Interpretation*. London: Butterworth Scientific Publications.

POPPER, K. R. [1967]: 'Quantum Mechanics without "The Observer" ', in Mario Bunge (*ed.*): *Quantum Theory and Reality*. Springer-Verlag, pp. 7–44.

ROBERTS, D. D. [1970]: 'On Peirce's Realism', *Transactions of the Charles S. Peirce Society*, **6**, pp. 67–83.

ROBIN, R. [1967]: *Annotated Catalogue of the Papers of Charles S. Peirce*. University of Massachusetts Press.

SKLAR, L. [1970]: 'Is Probability a Dispositional Property?', *Journal of Philosophy*, **67**, pp. 355–61.

THAYER, H. S. [1968]: *Meaning and Action*. Indianapolis: Bobbs-Merrill.

TURSMAN, R. [1970]: 'Will's Criticisms of Peirce's Frequency Theory of Probability', in R. Tursman (*ed.*): *Studies in Philosophy and in the History of Science*. Lawrence, Kansas: Cornado Press.

WILL, F. L. [1965]: 'The Preferability of Probable Beliefs', *Journal of Philosophy*, **62**.

Brit. J. Phil. Sci. 26 (1975), 133–141 *Printed in Great Britain*

Discussions

SOME NEW ASPECTS OF RELATIVITY: REMARKS ON KESWANI'S PAPER

In the early days of relativity, one of the facts given most emphasis was the invariance of Maxwell's equations with respect to the Lorentz transformation. Such a rich spectrum of results follows from the invariance of Maxwell's equations that it appeared as if the whole basis of the relativity theory was tied up with electrodynamics.

Maxwell's equations imply the wave equation

$$\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} - \Delta\phi = 0 \tag{1}$$

and hence also the equation

$$c^2 t^2 - x^2 - y^2 - z^2 = 0 \tag{2}$$

describing the geometry of spreading of the wave front.

Each transformation which leaves Maxwell's equations invariant leaves therefore also the equations (1) and (2) invariant. The equations (1) and (2), however, hold for a much wider range of physical phenomena than simply electrodynamic ones. The properties which the equations (1), (2) describe are, as is well known, the wave properties. Of course, the value of the constant c is different for various kinds of waves. It has one value for sound, another for the waves on the surface of water, and still another for electromagnetic waves.

Hence, it could be expected that the properties of Lorentz's type of transformation are associated with the wave properties in general rather than with the electromagnetic phenomena only. Actually the possibility of relativity in an acoustical world has already been discussed by Kar.[1]

However, we have to be careful about such considerations. Because Maxwell's equations do not follow from the wave equation (1), it is not guaranteed *a priori* that every transformation leaving the wave equation invariant leaves also the Maxwell equations invariant. So if we found the complete set of transformations leaving the wave equation invariant, we would also have to show which of them leave the Maxwell equations invariant.

With respect to the quite general axioms of one-to-one mapping, homogeneity of space and time,[2] cylindrical symmetry according to the x-axis[3] as well as some usual simplifying assumptions about the frames S and S', we start with the sets of linear equations.

$$\begin{aligned} x' &= g_{11}(x - vt) \\ y' &= g_{22}\, y \\ z' &= g_{22}\, z \\ t' &= g_{44}\, t + g_{41}\, x \end{aligned} \tag{3}$$

---

[1] Kar [1970].

[2] Navrátil and Podlaha [1966].

[3] Navrátil and Podlaha [1967].

$$\begin{aligned} x &= g^{+}_{11}\, x' + v\, g^{+}_{14} t \\ y &= g^{+}_{22}\, y' \\ z &= g^{+}_{22}\, z' \\ t &= g^{+}_{44}\, t' + g^{+}_{41}\, x' \end{aligned} \tag{4}$$

Keswani starts with the equations which are similar to the equations (3). Using further the equation (2) and the equation

$$c^2 t'^2 - x'^2 - y'^2 - z'^2 = 0 \tag{5}$$

he derives the Lorentz transformation.[1] However, at this point we must note something very important.

Generally, all the coefficients g_{ik}, g^{+}_{ik} *in the equations* (3), (4) are functions of v. Keswani's derivation loses its generality when he takes g_{ik} as constant factors independent of v. From the initial condition, $y' = y$ for $v = 0$ he derives $g_{22} = 1$ although in fact if we take g_{22} as a function of v as it really is, we cannot obtain $g_{22}(v) = 1$ merely from the initial conditions. It is important to remember this in deriving results from equations (3) and (4).

Substituting (3) into (5) we obtain

$$\begin{aligned} (c^2 g^2_{44} - v^2 g^2_{11})\, t^2 + (c^2 g^2_{41} - g^2_{11})\, x^2 - g^2_{22} y^2 & \\ - g^2_{22}\, z^2 + 2\,(c\, g_{44}\, g_{41} + v\, g^2_{11})\, x\, t &= 0 \end{aligned} \tag{6}$$

and comparing it with (2), gives us

$$\frac{c^2 g^2_{44} - v^2 g^2_{11}}{g^2_{22}} = c^2 \tag{7a}$$

$$\frac{c^2 g^2_{41} - g^2_{11}}{g^2_{22}} = -1 \tag{7b}$$

$$c\, g_{44}\, g_{41} + v\, g^2_{11} = 0. \tag{7c}$$

Let us further emphasise that we must not combine the equations (5) and (2) into the form $c^2 t'^2 - x'^2 - y'^2 - z'^2 = c^2 t^2 - x^2 - y^2 - z^2$.

From (7) it follows

$$g_{44} = g_{11} \tag{8a}$$

$$g_{41} = -g_{11}\, v/c^2 \tag{8b}$$

$$g_{11} = \frac{g_{22}}{\sqrt{(1 - v^2/c^2)}} \tag{8c}$$

From the equations (3) we have thus obtained

$$\begin{aligned} x' &= g_{22} \frac{x - v\, t}{\sqrt{(1 - v^2/c^2)}} \\ y' &= g_{22}\, y \\ z' &= g_{22}\, z \\ t' &= g_{22} \frac{t - v\, x/c^2}{\sqrt{(1 - v^2/c^2)}} \end{aligned} \tag{9}$$

---

[1] Keswani [1966].

In a similar way we obtain from the equations (4), using (2) and (5),

$$g^{+}_{41} = g^{+}_{11}\, v/c^2 \tag{10a}$$

$$g^{+}_{44} = g^{+}_{11} \tag{10b}$$

$$g^{+}_{11} = \frac{g^{+}_{22}}{\sqrt{(1-v^2/c^2)}} \tag{10c}$$

and hence

$$\left.\begin{aligned} x &= g^{+}_{22}\frac{x'+v\,t'}{\sqrt{(1-v^2/c^2)}}\\ y &= g^{+}_{22}\, y'\\ z &= g^{+}_{22}\, z'\\ t &= g^{+}_{22}\frac{t'+v\,x'/c^2}{\sqrt{(1-v^2/c^2)}} \end{aligned}\right\} \tag{11}$$

It is easy to see that from (9), (11) it follows

$$g^{+}_{22} = \frac{1}{g_{22}} \tag{12}$$

However, this is the only condition on the function g^{+}_{22} which follows from equations (9) and (11) and no condition is made on the function g_{22}. Whatever the function g_{22} is, the transformation (9) leaves the wave equation (2) invariant.

Keswani's derivation is thus generalised. Not only the Lorentz transformation but all transformations of the type (9) and (11) transform the spherical wave in S into a spherical wave in S'.

Now, it can be proved that the derived transformations (9) and (11) leave also the Maxwell equations invariant. This was proved by myself in detail in another paper.[1] Indeed, the transformations (9), (11) leave Maxwell's equations invariant and from this invariance no further conditions on the functions g_{22}, g^{+}_{22} follow.

The fact that g_{22} is some indeterminate function of v was already known to Lorentz[2] and Poincaré.[3] However, they took, as Einstein did later, $g_{22} = 1$. Poincaré gave as a reason for his choice that the transformations (9) form a group.[4] Einstein, as is well known, justified the choice $g_{22} = 1$ by the requirement of his principle of relativity that the frames S and S' have to be indistinguishable from each other. Keswani finds $g_{22} = 1$ from the initial condition $v = 0 \rightarrow y' = y$ by supposing g_{22} as a constant factor instead as a function of v.

We feel that the most natural way is to seek g_{22} on the basis of experiment.

As we have shown elsewhere,[5] from the analysis of the interferometrical experiments of the Thorndike–Kennedy type it follows that there exists a frame S° for which the function $g^{S^{\circ}}_{22}(v)$ has the form

$$g^{S^{\circ}}_{22}(v) = \frac{1}{(1-v^2/c^2)^{\frac{1}{2}+\gamma}}$$

where $\gamma \in \langle -1/2, 0\rangle$.

---

[1] Podlaha [1974].

[2] Lorentz [1904].

[3] Poincaré [1905].

[4] *Ibid.*

[5] Podlaha [1969].

The function $g_{22}^{S^\circ}(v)$ defines unambiguously the transformation $T^{S^\circ}(v)$ providing the transitions $S^\circ \rightarrow S(v)$, $S^\circ \rightarrow S(w), \ldots$, from the frame S° to the frames $S(v), S(w), \ldots$, moving with the velocities $v, w, \ldots$, with respect to S°, respectively.

Knowing this we can calculate the function $g_{22}^{S(v)}(u)$, estimating the transformation $T^{S(v)}(u)$ between the frames $S(v)$ and $S(w)$ where u is the velocity of the frame $S(w)$ with respect to the frame $S(v)$.

$$y^{S(v)} = g_{22}^{S^\circ}(v) \cdot y^{S^\circ} \tag{14a}$$

$$y^{S(w)} = g_{22}^{S^\circ}(w) \cdot y^{S^\circ} \tag{14b}$$

From (14*a*) and (14*b*) we get

$$y^{S(w)} = \frac{g_{22}^{S^\circ}(w)}{g_{22}^{S^\circ}(v)} \cdot y^{S(v)} \tag{15}$$

and hence

$$g_{22}^{S(v)}(u) = \frac{g_{22}^{S^\circ}(w)}{g_{22}^{S^\circ}(v)} = \left(\frac{1-v^2/c^2}{1-w^2/c^2}\right)^{1/2+\gamma} \tag{16}$$

whereby

$$w = \frac{v+u}{1+vu/c^2} \tag{17}$$

We see that to every couple of particular frames, S^i, S^k; S^p, S^q; ...; S^r, S^s corresponds just one coefficient g_{22}^{ik}; g_{22}^{pq}; ...; g_{22}^{rs}, and just one pair of velocities v_i, u_{ik}; $v'_p u_{pq}$; ..., defining unambiguously the transformations T^{ik}; T^{pq}; ...; T^{rs} for the transitions $S^i \rightarrow S^k$; $S^p \rightarrow S^q$; ...; $S^r \rightarrow S^s$ respectively. Further the transformation T^{ik} can be followed by the transformations T^{pq} and the transformation T^{pq} by the transformation T^{rs} if and only if the frame S^k is identical with the frame S^p and the frame S^q identical with the frame S^r. *If the frame S^k is not identical with the frame S^p, the transformation T^{pq} must not follow the transformation T^{ik}.* This holds analogously for the frames S^q; S^r and transformations T^{pq}; T^{rs}. This is because the transformations T^{ik}; T^{pq}; ...; T^{rs} are always bounded on the particular frame-couples S^i, S^k; S^p, S^q; ...; S^r, S^s respectively.

Now, it remains to determine the coefficient γ. This was done experimentally, first by Ives and Stilwell in 1938 by means of the transversal Doppler effect.[1] Ives and Stilwell found $\gamma = -1/2$ with an accuracy of about 10 per cent. The more recent measurements made by Mandelberg and Witten gave $\gamma = -1/2$ with an accuracy of 5 per cent.[2]

The Lorentz transformation is thus based on three experimental facts: (1) The wave properties of light, (2) The null results of the interferometrical experiments of the Thorndyke-Kennedy type, (3) The transversal Doppler effect. No other experiments are necessary. Especially let us note that the Trouton–Noble experiment is not able to contribute to the estimation of the space-time transformation at all.[3]

Einstein's principle of relativity was not necessary either. At this point our derivation again supports Keswani's conviction that, to build up relativity theory, it is not necessary to use Einstein's principle of relativity at all.[1]*

---

[1] Ives and Stillwell [1938].

[2] Mandelberg and Witten [1962].

[3] Podlaha [1974*a*], [1974*b*].

M. F. PODLAHA
Philosophisches Seminar der Georg-August-Universität, Göttingen

REFERENCES

IVES, H. E. and STILWELL, G. R. [1938]: 'An Experimental Study of the Rate of a Moving Atomic Clock', *Journal of the Optical Society of America*, **28**.

KAR, K. C. [1970]: 'Relativity in an Acoustical World', *Indian Journal of Theoretical Physics*, **18**.

KESWANI, G. H. [1965]: 'Origin and Concept of Relativity', *The British Journal for the Philosophy of Science*, **15**, pp. 286–306, and **16**, pp. 19–32 and 273–94.

LORENTZ, H. A. [1904]: 'Electromagnetic Phenomena in a System Moving with any Velocity Smaller than the Speed of Light', Proceedings of the Royal Academy of Amsterdam, **6**, pp. 809–34.

MANDELBERG, H. I. and WITTEN, L. [1962]: 'Experimental Verification of the Relativistic Doppler Effect', *Journal of the Optical Society of America*, **52**.

PODLAHA, M. [1969]: 'Lorentz Theory, Palacios Theory and Interferometrical Experiments', *Nuovo Cimento*, **64B**.

PODLAHA, M. [1974*a*]: 'Analysis of Invariance of Maxwell's Equations with Respect to Asymmetrical Linear Transformations of Voigt–Palacios–Gordon Type', *Indian Journal of Theoretical Physics*, **22**; see also note in *Bulletin de la Société Scientifique de Bretagne*, **48**, 248 (1973).

PODLAHA, M. [1974*b*]: 'The Axiomatic Foundations of the Theory of Special Relativity—Reply to Stiegler', *International Journal of Theoretical Physics*, **11**.

PODLAHA, M. and NAVRÁTIL, E. [1966]: 'On the Linearity of the Lorentz Transformation' *Acta Physica Austriaca*, **24**.

PODLAHA, M. and NAVRÁTIL, E. [1967]: 'Formulation of Axioms for Deriving the Lorentz Transformation, Palacios Transformation and Galilei Transformation', *Revista Real Academie Ci. Madrid*, **61**.

POINCARÉ, H. [1905]: 'Sur la Dynamique de l'Electron', *Comptes Rendus de l'Académie des Sciences*, **140**, pp. 1504–8.

MACH'S PRINCIPLE: MICRO- OR MACROPHYSICAL?

In his note in an earlier issue of this *Journal*,[2] Sachs raises several points that we cannot let pass without comment. We are pleased that he agrees that our previous examinations of Mach's Principle, 'get to the heart of current controversies on the subject of the Mach principle *vis à vis* the theory of general relativity.[3] His criticisms of our conjectures, however, are largely misleading or erroneous.

We shall first comment on Sachs's statements with regard to the Schwarzschild solution of the field equations contained in his point 3. Contrary to Sachs's implication that the Schwarzschild solution is merely an approximation, it should be noted that this solution is one of the few *exact* solutions of the field equations! Taking $R_{\mu\nu}=O$ (the exact free field equations) is entirely admissible in the case of a singularity (everywhere except at the singularity of course); and the solution of the field equations *exterior* to an extended, stationary, spherically symmetric source distribution will be this solution with a suitable choice of the mass for the singularity. The spherical symmetry of the solutions of the field equations is not 'imposed': it is a consequence of the implicit assumption of the homogeneity and isotropy of space-time (by the way, also an assumption implicit in Sachs's theory).

---

[1] Keswani [1966].

* I should like to thank Professor E. Scheibe for valuable discussions and for proposing a different symbolism which made the equations (13) to (17) and related paragraphs clearer and more comprehensive and compact than they had been in a previous draft.

[2] *Cf.* Sachs [1975].

[3] *Cf.* Woodward and Yourgrau [1972] and [1973].

As to the supposed asymptotic flatness of space-time at large distances from a Schwarzschild source, Sachs is, of course, correct when he points out that this is an independent boundary condition which is *not* entailed by the structure of the field equations.[1] But solutions of the field equations of general relativity cannot be obtained *without* such boundary conditions.[2] In fact, it was this point that led Einstein to conclude that he had failed to incorporate Mach's Principle in his theory!

Sachs states that, 'In accordance with the commonly accepted philosophy of science, if [we] wish to reject [his theory] in a scientific way, [we] must technically refute the mathematical derivations that [he has] published...'[3]. Frankly speaking, we find this claim astonishing. Whole hosts of formally consistent theories can be envisaged that have nothing whatsoever to do with reality. Must flaws in the mathematics be found in each *before* they can be rejected? Indeed, formal consistency is rarely the criterion by which a theory is evaluated. It is usually presumed that mathematical errors are absent. The criteria that are, or at any rate should be, applied are: (1) Is the conceptual substrate upon which the theory rests a plausible view of the structure of physical reality? (2) Do the predictions of the theory correspond to observation in all cases where the theory reasonably may be assumed to apply? Questions of formal consistency aside, Sachs's theory may be taken to task on at least the first of these criteria.

In this regard, let us point out that Sachs simply assumes Maxwellian electrodynamics, but with the 'implicit' rejection of all self-energy terms in the formalism.[4] Now this is manifestly unacceptable. Either one formulates a new electrodynamics in which these terms do not occur, or one explains physically why these terms do not give observable contributions. The problem of self-energy in field theory has occupied the minds of many of the most eminent physicists for the better part of a century. To merely 'implicitly' reject self-energy terms in a formalism can only be viewed as an instance of most flagrantly unconvincing arguments.

---

[1] Here it is of interest to note that the assumption of asymptotic flatness is something of a paradox, as Brillouin ([1970]) has pointed out. In general relativity it is routinely supposed that in the absence of matter the Minkowski metric of special relativity theory should be recovered. Yet this space-time is that in which the law of inertia (Newton's first law) is presumed to prevail, but inertial frames of reference can only be discerned by consideration of the motion of massive bodies *relative to others* whose existence is precluded by general relativity in this limit. Brillouin's observation that something is amiss with our concepts certainly seems justified. The problem is also unmistakably related to Mach's Principle.

[2] Sachs's claim that the flatness of space-time at an infinite distance from a single source is an approximation is misleading. At infinity space-time is *exactly* flat. In his theory, since there exists a homogeneous sea (aether) of electron-positron pairs, this is not the case.

[3] *Cf.* Sachs [1975].

[4] *Cf.* Sachs [1972], especially p. 252.

Sachs's claim that, 'the physical model of vacuum polarisation has never been shown to be unambiguously implied in a demonstrably consistent form, either logically or mathematically, by the computational techniques of renormalisation . . . .'[1] is highly misleading. Many physicists, beginning with Faraday,[2] have speculated that the vacuum is a polarisable entity, *but this conjecture is only empirically validated by experiments that confirm the predictions of quantum electrodynamics.* In quantum electrodynamics, the occurrence of vacuum polarisation is a logical and mathematical consequence of the self-energy renormalisation of the photon.[3] Notwithstanding that quantum electrodynamics has severe consistency problems, it should be observed that Sachs merely postulates *ad hoc* the existence of polarisable electron-positron pairs in the vacuum, while the existence of these pairs may be deduced from first principles in quantum electrodynamics. In a sense, Sachs's theory is but a more sophisticated version of the various dielectric theories of gravitation that have been proposed from time to time.[4]

These caveats, important as we feel them to be, are not our central criticism of Sachs's ideas. Our chief concern is that Sachs has distorted Mach's Principle in his theory to the point that it no longer contains the crucial and main physical import intended by Mach. Sachs has reduced Mach's Principle from a macro- to a microphysical principle by asserting that the inertia of objects is induced locally (within a distance of 10^{-15} *cm*) by interactions of matter (elementary particles) with a sea of electron-positron pairs that is supposed to permeate the vacuum. Sachs even has the temerity to maintain that this, in principle, was Mach's position! This is definitely not the case, as any patient perusal of Mach's analysis of the concept of absolute space shows beyond any doubt.[5] How could the inertia of local objects be causally related to the existence and distribution of the 'fixed stars'—as Mach contended—if the local vacuum (with its pairs) determines the inertia of objects, as Sachs maintains?[6]

In our opinion, the assertion that the inertial mass of an object should tend to zero as the remainder of the matter in the universe is removed is eminently plausible and, indeed, *required* by the principle of relativity. On the other hand, we find the notion that an object may be rendered inertialess solely by depleting the local vacuum of its electron-positron pairs frankly repulsive, for it is clearly at variance with the principle of global relativity as formulated by Mach.

We shall conclude with a few observations on Sachs's statement that he does not agree with our, 'rejection of the unified field approach, and therefore believe[s] that it is logically necessary to extend the Mach principle to all of the physical manifestations of matter.[1] We are not opposed, in principle, to the unified field approach. We do not accept, however, that the way in which this approach has been used in the past is fruitful. In our opinion, geometrisation is not the solution to the structure of matter. We do agree that Mach's Principle should be extended to all manifestations of matter. But this poses tremendous problems. For example, are we to assume, contrary to all accepted theory, that charge is induced by the presence of other charge in the universe? It may seem so. But this would imply that the magnitude of the charge of an elementary particle should be affected by the presence of electromagnetic fields. Yet no observations that would support the existence of such an effect are in evidence.

---

[1] *Cf.* Sachs [1975].

[2] Vacuum polarization was a feature of Faraday's theory of the structure of matter. *Cf.* Williams [1965], Agassi [1971], and a forthcoming monograph, *Faraday and Gravitation,* by one of the authors.

[3] *Cf.* Sakurai [1967], pp. 273–96.

[4] The most recent and highly developed of these theories, prior to that of Sachs, was constructed by Dicke ([1957]).

[5] *Cf.* Mach [1960], pp. 276–90.

[6] In fairness to Sachs, it must be admitted that he has incorporated some of the features implied by Mach's Principle. The inertia of matter is not inherent or totally self-induced, and the mass of a test particle goes to zero as the remaining matter in the cosmos (*including the electron-positron pairs in the vacuum*) is depleted.

Perhaps our inability to successfully assimilate Mach's Principle stems from a conceptual prejudice so deeply ingrained that we are hardly aware of it. A potential candidate for such a prejudice might be the bifurcation of physical reality into sources and fields (localised matter and its actions—carried by fields—on other matter). Our concept of sources, be they singularities of the fields or little extended objects, may be untenable. Faraday once entertained this idea.[2] The elimination of the concept of sources might be instrumental in solving the problem of self-energy as well as those attendant on Mach's Principle. General relativity, wherein source strengths depend on the presence and strength of the gravitational field, is a step in this direction, but evidently does not go far enough. The idea of writing down field equations that describe matter without sources admittedly seems bizarre, but many of our concepts today would have appeared bizarre to physicists as little as a hundred years ago. In any event, whatever the solution to the problem posed by Mach's Principle may be, it cannot coincide with Sachs's theory—for Mach's Principle is clearly a *macrophysical principle.*

JAMES F. WOODWARD
California State University, Fullerton
WOLFGANG YOURGRAU
University of Denver

REFERENCES

AGASSI, J. [1971]: *Faraday as a Natural Philosopher.*

BRILLOUIN, L. [1970]: *Relativity Reexamined.*

DICKE, R. [1957]: 'Gravitation without a Principle of Equivalence', *Reviews of Modern Physics*, **29**, pp. 363–76.

MACH, E. [1960]: *The Science of Mechanics.*

SACHS, M. [1972]: 'On the Electron-Muon Mass Doublet from General Relativity', *Nuovo Cimento*, **7B**, pp. 247–64.

SACHS, M. [1975]: 'On the Mach Principle and General Relativity', *British Journal for the Philosophy of Science*, **26**, pp. 49–51.

SAKURAI, J. J. [1967]: *Advanced Quantum Mechanics.*

WILLIAMS, L. P. [1965]: *Michael Faraday.*

---

[1] *Cf.* Sachs [1975].

[2] Faraday held this position during the 1840s, but relapsed into the sources and fields with substantial substrate concept in the early 1850s. *Cf.* the forthcoming monograph, *Faraday and Gravitation.*

Woodward, J. F. and Yourgrau, W. [1972]: 'The Incompatibility of Mach's Principle and the Principle of Equivalence in Current Gravitation Theory', *British Journal for the Philosophy of Science*, **23**, pp. 111–16.

Woodward, J. F. and Yourgrau, W. [1973]: 'Mach's Principle, The Equivalence Principle and Gravitation: A Rejoinder to Newburgh', *British Journal for the Philosophy of Science*, **24**, pp. 264–70.

Brit. J. Phil. Sci. 26 (1975), 143–149 *Printed in Great Britain*

Review Articles

THE END OF THE ROAD FOR INDUCTIVE LOGIC?*

This book was published shortly after the death of Carnap, the principal of its two author-editors. The first two papers in the book are due to him; Richard Jeffrey says in his Introduction that Carnap had almost finished revising the continuation of the second of them (which will appear in Volume II) at the time of his death. These papers of Carnap, in this and the next volume of these *Studies*, are therefore his final work in the field he entered with the *Logical Foundations of Probability*, continued with the *Continuum of Inductive Methods*, the *Induktive Logik und Wahrscheinlichkeit* (with Stegmüller), the 'Systematic Expositions' in Carnap's volume of Schilpp's *Library of Living Philosophers*, plus a host of papers in the subject. Earlier versions of the papers he contributed to this volume have, as a matter of fact, been in circulation for about a decade, but there have been changes at least of style in his contributions to this volume, which create an impression of novelty.

This volume is a collection: it contains the papers by Carnap already mentioned, and one each by Richard Jeffrey, Jürgen Humburg and Haim Gaifman. Jeffrey's Introduction states that *Studies in Inductive Logic and Probability* was planned already in 1960, as a 'non-periodical journal in which various authors would report work in progress towards an adequate system of inductive logic', which would take the place of the previously planned sequel to the *Logical Foundations* which never appeared in its intended form because of the new perspectives on the subject gained by Carnap and his school in the 1950's: that intended sequel, Jeffrey says, 'would be no mere elaboration of the outline at the end of the Logical Foundations, but the outcome of at least another decade's work, the shape of which could not yet be clearly seen.' Carnap's long paper 'A Basic System of Inductive Logic, 1' printed in this volume (Part 2 is to appear in the next) is the most detailed account of the work of this later period.

A version of Carnap's first paper in this book, here entitled 'Inductive Logic and Rational Decisions', was published in 1962 under the title 'The Aim of Inductive Logic'. For this reason I shall not dwell too long on its contents. It assumes a Bayesian model of decision-making, and argues for a construction of the probabilities involved in computing the expected utilities of acts as rational credibilities, if the preference ordering among acts induced by the expectations is to determine a rational decision. Learning from experience is presented as a form of Bayesian conditionalisation, each new observational datum being incorporated into the second argument of a suitable conditional probability function ('Credibility function'). Rational inductive behaviour is determined jointly by utilities and the contemporary posterior distribution.

None of this is novel Carnap, though some points are still worthy of comment.

---

[*] Review of Carnap, R. and Jeffrey, R. C. [1970]: *Studies in Inductive Logic and Probability*, University of California Press. Pp. 264.

First, a small point: nothing is said here about how, or whether, this model can be applied in the field of theory selection. Section 7 of this paper is quite explicit that an account of inductive reasoning permitting the *acceptance* of statements about the future, including general hypotheses, has overreached itself, because of Hume's well-known objections; but those objections are decisive only if 'acceptance' means 'acceptance as true'. Nothing is said here about the possibility of a decision-theoretic account of theory selection in obedience to a rule of maximising what Hempel called 'epistemic utilities'.

Second, as Carnap points out, only the uninterpreted formulas of his systems are 'logical'; what he presents here is not a piece of logic but a controversial theory of rationality. In particular, there is nothing particularly logical about logical probability in Carnap's sense of the term. Carnap's reason for bestowing a logical status on his various probability functions is that their class of defining conditions refers to no contingent fact (p. 26), which would seem to imply that length, area, volume and probability are all part of logic, and that we can speak of logical length, *etc.*

Finally, it is by no means clear that the notion of probability 'adequate in the context of [normative] decision theory', should invariably be that of rational degree of belief, as Carnap maintains. Where the outcomes of an act depend on a set of exclusive and exhaustive possibilities over which a physical probability distribution exists, the most reasonable thing to do would seem to be to use one's knowledge of this distribution in computing the desired expectations. To our objection stated thus there is a plausible and familiar defence: namely that while this is certainly true, we can never by observation estimate with an arbitrarily high guaranteed accuracy any statistical parameter, and consequently must rely on a (subjective) conditional probability distribution for the parameter, and use this distribution itself, integrating over all values of the parameter, to give the required (subjective) probabilities of the outcome at some future trial, conditional on the observed sample. Hence, while it may be conceded that a recognition of physical distributions, where they exist, is indispensable, it might plausibly be maintained that a correct understanding of their properties is obtained only in the context of a personal probability theory. This may or may not be the case, but even if it were it would not entirely meet the objection to Carnap. For Carnap's credibility functions are all symmetric, that is to say, such that the probability assigned to a finite sequence of elementary events does not depend on the order of the events in that sequence; and one would normally use such a measure only when one had reason to believe that the events in question did not exhibit causal dependencies. A natural question to ask at this stage is whether it is possible, within the framework of a symmetric credibility function, to discuss and acquire from appropriate sample evidence, knowledge about, for example, stochastic processes in which there *are* definite dependencies. This is a technical question about the properties of the sorts of probability spaces admissible within Carnap's theory, and this is the sort of question, therefore, that should be answered by an examination of his second paper in this volume, where some of the formal details of the theory (the rest are assigned to part II) are spelt out.

This second paper expounds a theory less constrained formally by a particular language than the systems in the earlier λ-continuum, for example. The measures are now defined in a standard way on σ-fields of events, here perversely called propositions. Characteristically the sample spaces represent all distributions of attributes from some family among a denumerable set of indices (pp. 43, 45). The σ-fields are generated by the subsets of these spaces definable by the quantifier-free formulas of a monadic language (p. 59), *i.e.* by the cylinder sets corresponding to these formulas. But for the restriction to a denumerable set of attributes (p. 46) there would be nothing to distinguish Carnap's theory formally from the standard theory of probability measures on product spaces. The distinctively Carnapian character of this theory is therefore largely restricted to interpretation. In his first paper Carnap said that this is a global theory of rationality, its global character deriving from: (*i*) the gradual extension of the credibility function over enlargements of some initial field, determined by the extension of some primitive language, (*ii*) the fact that the credibility function (or rather functions: we are still far from uniqueness) is determined independently of any experience, and (*iii*) the invariable rationality conditions of regularity (p. 101) and symmetry (p. 117) imposed on the credibility functions. These features differentiate Carnap's from the subjectivistic theory of Ramsey, for example, which is a more local affair in which experience is permitted a role in determining both the individual's pre-experimental beliefs and the appropriate method of classifying the elements of a problem. This last consideration, that experience plays characteristically a much more than formal role in determining the classification of data, is important in any theory of the growth of knowledge. In Carnap's theory, experience is allowed to provide substantive information only after it has been classified within a given language with which is associated a class of Carnap's credibility functions. Experience cannot, consequently, influence the choice of language according to any criterion of whether a particular combination of language and credibility functions permits an accurate determination of a particular state of affairs. This sort of objection is endemic to global theories, and seems quite fatal to Carnap's. We can illustrate this general point with the problem raised above of how one could learn that there is dependence and how one could estimate its magnitude within the framework of a symmetric probability measure which apparently denies it. In fact, it is easy to show that it is possible to use Carnap's theory to obtain an accurate estimate of the extent of dependence; but it is also possible to show that a no less legitimate use of Carnap's theory, judged according to the canons supplied by that theory, yields a completely unrealistic estimate: it all depends on the way the problem is described. Suppose we are dealing with a Markov chain. Then by considering a sample space of sequences of sample functions we do not implicitly deny the existence of dependence; we simply assume that the order in which the sample functions are taken is irrelevant, which may well be the case. We can now, with the symmetric credibility measure, estimate from the appropriate frequencies the transition probabilities of the chain. Perhaps the simplest method of doing this is to consider the subspace S of a space of denumerable sequences of pairs of outcomes, whose first members are all a given state i and define a sequence of random variables x_n on the subspace such that if in the nth pair i is followed by j then $x_n = 1$; otherwise $x_n = 0$. The transition probability p_{ij} is therefore equal to $P(x_n = 1)/P(\mathrm{S}) = P(x_m = 1)/P(S)$, where the probabilities are of course understood in a physical sense, and we can use an integral representation theorem of de Finetti's about symmetric measures to provide a posterior distribution for a variable which we can identify with the physical binomial parameter $p_{ij} = P(x=1)/P(S)$. We can now discuss also the credibility (in the sense of the measure we have hypothetically adopted) of state j following state i, given some sample evidence. Suppose we have observed an initial segment f_m of a sample function f. If we partition f_m sequentially in pairs and throw out those whose first member is not i, we are left with a k-tuple which can be represented under the mappings x_n, as a k-fold sequence of 0 and 1; let us call this E_k and suppose there are r 1's in E_k. So we wish to evaluate

$$P(x = 1/E_k)$$

and this can be done by applying the formula for total probability, integrating p_{ij} throughout its range to give

$$\frac{\int_0^1 z^{r+1}(1-z)^{k-r}\mathrm{d}F(z)}{\int_0^1 z^r(1-z)^{k-r}\mathrm{d}F(z)}$$

F(z) is a distribution of function such that F(0) = 0, and F(z) = 1 for $z>1$. This is all very well, but if we adopt a different method of presenting the problem and its data we will actually get a different result. If we take the sample space to consist of sample functions and consider the probability of state j occurring given that state i has just occurred and before it the sequence of states represented by f_m, we shall get an answer based on the supposition that we have a straightforward case of multinominal sampling, where there is no physical dependence between successive states. This little example shows that depending on the way a problem is described, answers to it will characteristically differ; and in this case on answer is manifestly correct and one incorrect.

A case of a description-dependent result not clearly more or less preferable to another is the notorious Rule of Succession. We get this result for events which are simple alternatives if we adopt the limiting form of a symmetric credibility measure which assigns equal prior weights to the possible proportions of the alternatives in a hypothetical urn: the theory of this measure was fairly thoroughly investigated by Laplace, and re-discovered by Carnap as m^*. Now suppose we are throwing a die, and we are interested whether a six will occur at the $m+1$ th. If we define a sequence of random variables on sequences of throws that take the value 1 for a six at the ith place, and 0 otherwise, the solution is given by the Rule of Succession: the required probability, or credibility is $r+1/m+2$ where r sixes were noted among the previous m throws. If we consider the problem as one of multi as opposed to binomial sampling, on the other hand, and represent the observed results just as they occurred, for example, we shall not get the Rule of Succession, but its analogue for the multinomial case, where the probability in question is $r+1/m+6$.

Which is the correct answer? I do not believe that there is a right answer *i.e.* correct odds on six at the next throw, given the evidence, nor is the one set of odds clearly superior to the other.

This consideration by itself casts doubt on the adequacy of Carnap's theory sketched in the two papers in this volume, for there is presumably only one degree of credibility in any determinate situation. Two available options are: first to concede the field to the personalists, if one wishes to retain as much of this subjective probability approach as one can in the analysis of learning from experience; or else to abandon their approach altogether and use the theory of decision-taking under conditions of risk just when a physical distribution exists.

Finally, the reader should be warned that Carnap does not make life easy for him. In the second paper of this volume, for example, Carnap's quest for precision seems to me to make inordinately severe demands upon the reader; but it is not clear that this austerity, more extreme, if anything, than that displayed in the corresponding sections of his earlier *Logical Foundations of Probability*, is always justified by reason either of rigour or clarity. Some, in other words, of Carnap's formality seems to me to be quite gratuitous, and inimical to an easy understanding of his work. There is another aspect of Carnap's papers here that I personally found rather irritating; and that is the fact that they convey the bland impression that there is nothing in the least controversial about a theory which has been in one form or another, a lightning-conductor for destructive criticism for a century or more.[1] After these few remarks, which I hope are not too ungracious, let us move on to the other papers in this volume.

The last two in the volume, by Gaifman and Humburg, are concerned with eliciting properties of symmetric measures subject to further constraints, with the aid of the theorem of de Finetti mentioned earlier. What is shown is, first of all, that the following 'principle of non-negative relevance' is satisfied by all symmetric measures $\mathscr{C}$ on (infinite) product spaces:

$$\mathscr{C}(Ma|E \cap Mb) > \mathscr{C}(Ma|E)$$

where E is a singular statement not mentioning a or b, while if the formula ('Reichenbach's axiom'):

$$\lim_{n \to \infty}[\, \mathscr{C}(Q_j b_{n+1}|E_n) - s_j(n)/n] = 0$$

where E_n reports $s_j(n)$ occurrences of Q_j in a sample of size n, is appended, then the following recursion formula ('positive instantial relevance') is provable:

$$\mathscr{C}(Ma|E \cap Mb) > \mathscr{C}(Ma|E)$$

Carnap maintains that this principle expresses 'one of the basic characteristics of customary inductive reasoning' (p. 261).

It is interesting to ponder, briefly, the status of the principle of positive instantial relevance. It allegedly represents a type of learning from experience—specifically, of representing changes in one's measured expectations concerning the occurrence of a future event, stimulated by accumulating factual data. A plausible rival view is that we assess the reliability of predictions according to whether they follow from an accepted body of theory and a set of accepted initial conditions; but an objection to this view, as stated, is that a given prediction in general follows from a number of mutually exclusive theories, each of which stands in a determinate relation of being, more or less, supported by accumulated empirical data, and we must therefore anyway consider not one, but many mediating theories; and consider their support by these data. Furthermore, for an empiricist, mention of these theories should strictly be eliminable in assessing the support of predictions by data, for the ultimate justifying factors are the data themselves, which should stand in a determinate relation of support to any given prediction. In other words, an empiricist will make an 'independence of path' assumption, that the credence which ought to be granted to Pa, given E, depends only on Pa and E. Granted the assumption that conditional 'rational' credence is representable as a conditional probability function, the mediating role of theories is explained as follows. Let $\mathrm{T}_1, \ldots, \mathrm{T}_n$ be mutually exclusive theories in some language, each of which entails a singular prediction $\mathrm{Q}(a)$ and let $S = \neg\,(T_1 \vee \ldots \vee T_n)$. Then

---

[1] See Lakatos [1968] for some detailed criticisms of Carnap's own programme.

$$P(Q(a)|E) = \Sigma_i P(Q(a) \wedge T_i|E) + P(Q(a) \wedge S|E)$$
$$= \Sigma_i P(T_i|E) + P(S'|E)$$

where S' is the set of strongest sentences whose disjunction is equivalent to $Q(a)$ (to split S up in this way requires that our universe of individuals be finite). Thus the credence which should be assigned to $Q(a)$, on E, is just the sum of the degrees of credence, given E, assigned to the mediating 'rival' theories, on E, and S on E.

Carnap's implied thesis, that each new instance of a predicate should have a positive reinforcing effect on our confidence in a corresponding prediction is, at the least, arguable (though he himself provides no argument for it). If this is taken to be a characteristic feature of the processes of belief-formation of the human species, I should like to see some empirical evidence for it. To voice my own opinion, such an empirical investigation would be of far greater informative value than the sorts of idealised ('rational') belief systems investigated *a priori* by Carnap and his school.

The remaining paper in the volume is Richard Jeffrey's 'Probability Measures and Integrals'. This is a proof-free account of the notions of measure, measurability, measurable function and integration with respect to a measure. I can't think why on earth it was written, let alone included in this volume, for there are legion textbooks of varying difficulty dealing with these notions (and also supplying proofs) to which, indeed Jeffrey gives elaborate reference. Nor is this rather gratuitous exposition always very clear: I defy anybody who does not understand the concept 'conditioning with respect to a σ-field' to be able to grasp this notion as presented by Jeffrey.

The author/editors of this volume contend that what they have presented here 'is to be viewed as a foundation, not as a completed structure'. But the foundation does not appear to me to be at all firmly established. I have already pointed to what I regard as one very serious difficulty encountered by the theory, and that is its apparent placing of the revision of the system of categories within which learning from experience is done, and even the mode of application of that system to given problems, beyond the limits of rational discussion. This is one aspect of a more general criticism of the Bayesian approach, that it construes rational learning too narrowly. Another aspect is represented by the fact that the assignments of conditional probabilities are not themselves subject to rational revision, unless they fail to satisfy the rather general requirements of coherence (*i.e.* considered as the property of a set of conditional betting-quotients consisting in their being immune to a Dutch Book being made against them), symmetry, regularity, and to judge by what Carnap says, the principle of positive instantial relevance. Mention of this last brings me to a second type of criticism of the Carnapian programme, and one raised sharply by Sir Karl Popper and his school, namely concerning the types of thing actually decreed to be rational *within* that programme. That the principle of positive instantial relevance, for example, defines rational changes of belief in a prediction, given accumulating evidence consisting of positive and negative instances of the predicate involved in the prediction, is very much open to question.

The principle of symmetry, to take another example, which is a vestigial trace of the notorious Principle of Indifference, is defended by Carnap with an argument by analogy, but an exceedingly poor one, it seems to me. The argument is that the validity of deductive inferences is invariant under permutation of the individual constants appearing in premisses and conclusion, and that therefore it seems plausible that the same invariance should characterise inductively valid inferences. But this is surely a *non sequitur*: the invariance in in the case of deductive inferences is a simple corollary of the semantic property of deductive inferences of preserving truth, a property not of course possessed by non-deductive inferences.

There are other examples of criticism of the rationality principles adopted by Carnap—for example that produced by Hacking,[1] casting doubt on the Bayesian claim that rational changes of belief through time are necessarily representable by a conditional probability function. Had some of this criticism, coming from people whose competence is to be respected, been discussed or even mentioned in this volume, the editors would have done a great service to the novice in this field. As it is they seem vulnerable to the very serious charge of wilfully misleading him.

COLIN HOWSON
London School of Economics

REFERENCES

HACKING, I. [1967]: 'Slightly more realistic personal probability', *Philosophy of Science*, 34.

LAKATOS, I. [1968]: 'Changes in the problem of inductive logic', in I. Lakatos (*ed.*): *The Problem of Inductive Logic.*

[1] See Hacking [1967].

Brit. *J. Phil. Sci.* 26 (1975), 151–163 *Printed in Great Britain*

CREAKING SEMAPHORE AND BEYOND: A CONSIDERATION OF SHACKLE'S 'EPISTEMICS AND ECONOMICS'*

If we attempt to understand economic life by supposing that in it men apply reason to their circumstances, the question naturally arises: what *can* men know about their circumstances? Or even: what *do* men know about their circumstances? Having answered these questions we could then go on to ask: what happens when men apply reason to this knowledge? In fact, economic theory has not proceeded in this manner, but the other way round. Instead of asking how reason can be applied to the knowledge that men can or do have of their economic circumstances, it asks how reason can be applied to circumstances which are perfectly known. The problems of what can be known and how it can come to be known—problems of ignorance, uncertainty, risk, deception, delusion, perception, conjecture, adaptation and learning—are then tackled as a complication or refinement on the theory. Is this a satisfactory procedure? Are there economic phenomena which depend centrally and irreducibly on that which is not or cannot be known? Or is knowledge-deficiency a complication on the basic theory which can be introduced later as a refinement?

The existing mainstream of economic theory has developed in a manner which has accommodated these knowledge-deficiency problems as refinements to the theory of economic action rather than rudiments of it. This is not to deny that the mainstream has seen these problems as important and has devoted great efforts to their analysis. G. L. S. Shackle, however, rejects the results of these efforts. His reasons for doing so are set out in this book.

From another angle, this book is about the roles of reason and time in human affairs. Shackle's thesis is that the two are fundamentally antagonistic: that any attempt to accommodate both of them into a scheme of things leads to inconsistencies. Why should this be so? The application of reason to circumstances can only be made watertight and decisive if we exclude the novelty, uncertainty and ignorance that the future holds; for reason to function to its full capabilities, we must confine ourselves to a scheme in which time is either excluded altogether (as discussed in Book IV) or in which it appears in only an artificial way (as discussed in Book V). But what if an accommodation of real (novelty and uncertainty-bearing) time *and* the reasoning faculty of the economic actor are both essential to our understanding of economic life? If these two ideas are fundamentally antagonistic to one another, it follows that any attempt to capture one of them in a formal scheme will tend to exclude the other. Accordingly, a formalisation which makes watertight the operation of reason on the part of the economic actor must exclude the elapsing of time; and in the construction of any formalism one is faced with a choice as to which of these two ideas will be captured with decisive explicitness and which will be excluded in order to make

---

* Review of G. L. S. Shackle [1972]: *Epistemics and Economics: a Critique of Economic Doctrines*. Cambridge: Cambridge University Press. £7.80. Pp. vii+482.

The author would like to acknowledge the helpful comments and criticisms he has received from C. Bowe, S. J. Latsis and J. R. Ravetz, none of whom should be identified with the views expressed here, nor held responsible for the remaining defects of exposition.

this possible. Shackle, however, recognises a third possibility, which is to embrace a mode of formulation which rejects formalisation as the ideal of clarity and rigour. This leads him into direct conflict with the contemporary emphasis on formalising and axiomatic methods, and the general high esteem in which the deductive mode of reasoning is held. I shall argue in my later discussion, however, that the mainstream commitment to strict logic is more apparent than real: that it abandons strict logic in a quite different, much more clandestine, but no less decisive way than Shackle does.

The associated theme that the essential components of economic understanding can be held together only by the deliberate and careful cultivation of a language for imprecision leads Shackle into conflict with the orthodox view which sees the theory of probability as the vehicle whereby problems of knowledge-deficiency and knowledge-surrogates, both in the economic agent and the economist, are to be dealt with. This view Shackle sees as a misguided attempt to assimilate thought to arithmetic. Shackle's position also has major and radical implications for the interpretation of the recent history of economic thought, most notably for the significance we attach to the work of Marshall and Keynes, as we shall see.

'The simplest way to apply reason to the task of understanding conduct was to assume that conduct itself was based on reason'.[1] Thus, in economics, 'The theoretician assumes that men pursue their interests by applying reason to their circumstances. He assumes that they are both rational and fully informed'.[2] But reason cannot be applied directly to circumstances, but only to the knowledge (or its substitute) that men have of their circumstances. And circumstances do not establish knowledge: knowledge in the form in which reason may be applied to it is the creation of the human mind. There is more to conduct than passively 'recording' circumstances and then, as an essentially computational matter, acting in these recorded circumstances in accordance with ones interests. From the world of situations, events and processes to that of statements and inferences there is a conceptual gulf, the bridging of which requires that the mental processes of the actor be both active and creative. It is the investigation of the process of bridging this unavoidable gap that Shackle is designating by the term 'Epistemics'. But, in the light of this perspective, what sense can we give to the assumption that each agent is 'fully informed'? The final answer given by Shackle will be: none whatsoever, it is a completely incoherent idea.

Shackle's discussion may be put into sharper focus if we make central a distinction which Shackle himself touches on in passing but does not exploit in the organisation of his arguments: it is the distinction between what I have already referred to as the existence of *knowledge-deficiency* and the provision of *knowledge-surrogates* (these are not Shackle's terms). We may then, with the aid of this terminology, say that the subject of epistemics (and what is ordinarily covered in economics by the blanket term 'uncertainty') has two aspects or faces.

First, there is the recognition, delimitation of the extent, and characterisation of the types of *knowledge-deficiency*. This covers such things as risk, uncertainty, mistakenness, ignorance, deception and delusion. Second, there is the recognition, and characterisation of the different types of *knowledge-surrogate*: conjecture, expectation, perception, learning, adaptation and so on. 'Knowledge' itself does not appear on the list, for we can have knowledge only of the past; the reasoning which underlies conduct, relates to future, as yet non-existent, states of affairs. The reason why this distinction between knowledge-deficiency and knowledge surrogate has never been clear within economics is perhaps a result of the influence of the application of the theory of probability, which conflates both issues. As Shackle remarks:

---

[1] Shackle [1972], p. 148.

[2] *Ibid.*, p. 4.

> *Probability* is thus a word with two quite opposed uses. It is at the same time the name of a kind of knowledge, and an admission of a lack of knowledge. Because of its reference to these two highly contrasting and opposed ideas, the word probability has been seized on as an incantation to perform what reason declares impossible, the prescription of rational conduct in face of ignorance concerning the outcome of rival courses of conduct.[1]

Although most of the economics profession takes it for granted that the probability calculus performs the task that epistemics is here put to, Shackle's view would be that the muddling together of the two quite distinct ideas that the probability calculus involves results in its being adequate for the discussion of neither. We shall return to the discussion of probability later.

To summarise so far: we have arrived at the view that the gap between circumstances and conduct cannot be bridged by reason alone (even given the clearly-specified interests of the agent) nor by its handmaiden, logic; it requires constructive and creative thought. And all this makes no mention of the influence of time, the effect of which, Shackle argues at length, is to reinforce the views already reached.

'. . . the theoretician is confronted with a stark choice. He can reject rationality or time.[2] The rationality here referred to is, of course, that supposedly underlying conduct, not the understanding of conduct: it is that of the economic agent not his investigator. The reasons for this stark choice are intimately connected with the fact of knowledge-deficiency and the process of attaining and improving knowledge-surrogates on the part of the economic agent, as we shall see.

Each agent's conduct is based, inescapably, on his expectations (knowledge-surrogate) about his circumstances. The circumstances that are relevant in determining the consequences of alternative possible courses of conduct are those beginning at the present moment and extending into the future, that is, circumstances which are non-existent at the time of decision. But these future circumstances are in part the product of the conduct of others ('the state of the market', for example, will be the result of all the decisions, including that of the agent in question). To be 'fully informed' about the consequences of one's own decision, one would have to know what everyone else is deciding *at the same time*.[2] But this is equally true of each of the other agents: each must know what the others have decided before he himself can act in a 'fully informed' way. There is thus a simultaneity problem at the root of the problem of knowledge-deficiency in a social context. Indeed, it can be seen that, in a world where agents influence one another via their conduct, the fact that, at any moment, one agent is 'fully informed' precludes any other within the net of interdependence from also being so. (Or, more ominously, if the net of interdependence encompasses all, there cannot be more than one omniscient being).[1] Agents, can, however, experience the appearance or illusion of being 'fully informed' if their choices have been, somehow, *pre-reconciled*, so that the simultaneity problem is circumvented (or, as far as understanding is concerned, evaded). In this case the conduct of each agent is based on expectations which are, in the aggregate, 'self-fulfilling' rather than 'correct'. It is in the situation of 'equilibrium' that such a state of affairs occurs, and the epistemics of the process become straightforward. For in full equilibrium, market prices become perfect knowledge-surrogates. This is an interesting result, and the merits that are claimed for pricing systems and market processes stem from precisely this idea. But how do we get into this pre-reconciled state of full equilibrium? Is there any presumption that we are typically anywhere near it, or tending towards it? How would we recognise it if we were in it? For if we are not in it, the epistemic problem is not solved so simply as by invoking market prices alone.[2] The problem of how equilibrium could be reached is a troublesome one, and has been approached in various ways: Walras thought of it in terms of hypothetical *tâtonnements*; Edgeworth in terms of a hypothetical process of 'recontracting'. But in a world where out-of-equilibrium trading is binding rather than hypothetical, there is no presumption that market prices are very serviceable as knowledge surrogates; indeed, they are generally misinformation-surrogates.[3]

---

[1] *Ibid.*, pp. 15–16.

[2] *Ibid.*, p. xvi.

Shackle contrasts the method in which epistemic problems become straightforward—which he calls the analytic method—with what he calls constructive theories, in which time and problems of knowledge-deficiency play a key part. Thus:

> For . . . in the nature of the human imprisonment in time, constructive theories must describe a process of groping experiment or gambling for knowledge, of being wrong many times in order to be ultimately right,[4] while analytic theories describe what things will be like when, if ever, they have attained rightness.[5]
>
> The analytic method may thus produce an answer which is unique and unarguable to the question: How would things be in a momentary world of perfect pre-reconciled choice?[6]

---

[1] There may be none.

[2] In fact the result that in equilibrium market prices are perfect knowledge-surrogates is something of a swindle or, at best, a piece of conceptual conjuring. This is so because all the epistemic problems have to be solved in *reaching* equilibrium (*e.g.* by the Walrasian auctioneer). The reason that market prices 'reflect' everything that traders need to know about the markets is because—somehow—they have been rigged to do so. But if the individual trader has no way of knowing whether a state of general equilibrium obtains or not, what use to him is the knowledge that if it did obtain, all the market information he need consider would be contained in market prices?

[3] Indeed, such trading at disequilibrium prices and the associated transmission of misinformation has been put forward as the key to the interpretation of 'Keynesian Economics' and to the whole of monetary and employment theory. See Leijonhufvud [1968].

[4] Shackle here concedes too much, for there is no presumption, especially under changing conditions, that the agent will be 'ultimately right'.

[5] *Ibid.*, p. 53.

[6] *Ibid.*, p. 54.

Of theories of speculation, in particular, he writes:

> Analysts who were candid with themselves and their readers have found it necessary to exclude expectation from their analysis in order to make it rigorous. Weather forecasting, also, would be easier in the absence of an atmosphere.[1]

All of these difficulties are reinforced by the introduction of time; how, then, are we to proceed with its incorporation into the theory of economic processes? How, in particular are we to reconcile it with the rôle that reason has been chosen to play in our conception of the agent's decision about conduct; for in reinforcing these difficulties, we undermine the possibilities for the application of reason by the agent.

The standard way in which time is introduced into economic theory is by having models with time lags in the equations. Shackle quickly disposes of this by pointing out that it *poses* a question rather than answers one:

> Such a scheme must lead us to ask what vehicle conveys from the situations or events at one date the influence it is supposed to exert on those of another? *How do things grow out of each other through time?*[2]

One way out of these difficulties is simply to abandon time; to confine one's attention to the 'momentary world of perfect pre-reconciled choice'.[3] This possibility, the method of 'statics', is discussed at length by Shackle in Book IV and also in connection with the interpretation of Keynes's *General Theory* (in chs. 16, 17 and 20). In the method of statics, one might expect, there is some hope of keeping things sufficiently watertight for the agent's reason to have something definite to work on: the impoverishment of the framework of thought by the exclusion of time may be compensated for by the rigour with which one can formulate the now self-contained static theory. I shall argue later, however, that this rigour is only apparent and dissolves when one comes to scrutinise the way in which such theories are put to explanatory or predictive use via the method of comparative statics.

The second possibility is what I will call 'the de-naturing of time'. Thus, one can consider agents choosing not a self-contained, momentary act of conduct, but rather a whole time-path or plan for future conduct. But this brings us immediately up against the previously discussed simultaneity problem. And this, in turn, means that the theory can only be made workable by supposing that these plans or time paths have somehow been pre-reconciled so that they are all mutually consistent and realisable. Accordingly, in such a situation, the expectations on which plans were based could all be fulfilled (or, more precisely, would be 'self-generating'). This would then mean that the future produced no surprises: 'time' elapses but bears no novelty and no uncertainty. It is a future about which all agents have the perfect knowledge-surrogates for achieving the appearance of being 'fully informed'. It will be no surprise to learn that in such a situation market prices (including now certain inter-temporal prices) are perfect knowledge-surrogates, and the previous discussion in the static context carries over *mutatis mutandis*.

---

[1] *Ibid.*, p. 231.

[2] *Ibid.*, p. 58, emphasis added.

[3] *Ibid.*, p. 54 and above.
Cf. 'In their models [those of the 'latter day neo-classicals'] it is explicitly assumed that there is and has always been correct foresight, or else "capital" is malleable so that the past can be undone (without cost) and brought into equilibrium with the future; in short, they abolish time.' Robinson [1971*a*], p. 89.

There are other ways in which time may be de-natured for the purpose of its nominal introduction into the scheme of economic theory, and I cannot here hope to do justice to Shackle's discussion of these in Book V. Lack of space precludes consideration of Shackle's discussion of the assimilation of time to space and, in the same context, of what we will have to call the 'epistemic status' of business cycle theories (or 'business cycle engines' as he calls them). It is worth mentioning, in passing, the light which is cast on the interpretation of Marshall's *Principles* by all this. Shackle writes:

> The real character of Marshall's *Principles* seems to me to go almost unremarked by his commentators. Marshall sought to *accommodate* the intractable implications of time in an analysis appealing essentially to logic. He has been accused of lack of rigour. How else, except by a relaxation of rigour, can a compromise between incompatibles be achieved?[1]

Marshall himself wrote:

> Fragmentary statistical hypotheses are used as temporary auxiliaries to dynamical—or rather biological—conceptions, but the central idea of economics, even when its Foundations alone are under discussion, must be that of living force and movement.[2]

Shackle's remarks on the interpretation of Marshall's *Principles* bring us neatly to the third possibility of the accommodation of time within economic theory and its reconciliation with the rôle that reason is to play in it. This is the possibility advocated by Shackle himself: the abandonment of formalisation as an ideal. At its most basic, Shackle's case for this is simply that both reason and time are essential for an adequate conception of the economic process, and this is the only way in which they can both be accommodated, however precariously, without serious conceptual distortion to one or other of them. At another level, however, he shows how, in precisely those ways in which their work was fruitful and insightful, Marshall and Keynes made just this sacrifice of formalisation in favour of the accommodation of time and what it involves.

What Shackle does not consider is the point that, when it comes to being put to some use, the static method abandons its own formalisation anyway. The choice then becomes less dramatic: between abandoning formalisation openly or abandoning it in a surreptitious way.[3]

In strict logic, a comparative static result consists of showing that, if one of the underlying conditions were to be different from what it is in a particular period, the economic adaptation to those conditions would differ from what it would otherwise have been. A comparative static result is a comparison between mutually exclusive possibilities. The question now arises as to whether such a result can explain anything. Could it, perhaps, explain the change in some aspects of economic configuration in terms of the underlying conditions to which this configuration is an adaptation? The answer is, very roughly, yes, but within the confines of a strict formalisation, no. For an explanation of how one change leads to another change would have to give some sort of narrative (or, less guardedly, causal narrative) of how one change does indeed *produce* the other. And the comparative static method is incapable of this, by its nature. Indeed, by its nature, the method can say nothing about 'what the result of a change will be' much less 'why', for it is not addressed to the analysis of change: it is addressed to the analysis of perfect adaptation. Therefore it follows that the use of the comparative static method as an explanatory device involves a blatant violation of the strict logic of the static conception: it involves using a conception addressed wholly to situations in which nothing need change ostensibly to explain why something has changed. The static conception might just *accommodate* change, but it cannot explain it.[1]

---

[1] *Ibid.*, p. xvii.

[2] Marshall [1920], pp. xiv–xv.

[3] In what follows I shall draw freely on my discussion of explanation and prediction in Coddington [1972] without making detailed citations.

But what of prediction? Perhaps the comparative static method can be put to that use. The answer will again be, very roughly, it can, but strictly, it cannot.

Now a prediction takes the form: 'If something happens, then something else happens', without being committed to saying anything about *why* the one follows or is associated with the other. But to deploy the comparative static method in this way, the 'something happens' would have to be a change from one equilibrium to another (for nothing 'happens' in equilibrium). But, as we have seen, static equilibria are mutually exclusive possibilities; a 'change' from one to another does not correspond to a 'change' in the economic world ('something happens') but to a change in the set of presuppositions we are making. A comparative static result can therefore only be transmuted into a prediction rule (or an explanation) by fudging the strict logic of the formalism: by confusing a *synchronous difference* in presupposition with a *temporal change* in conditions.

None of this should be surprising, for it is simply saying that to explain a change one needs a conception of why change occurs (not a conception of the circumstances under which no change is called for); and to predict the results of changed conditions, one needs a model of the effects of changing conditions (not of the perfect adaptation of effects to a particular set of conditions). Nor should any of this be taken as an attempt to denigrate the comparative static method, but rather to show just how self-denying a genuine commitment to the strict logic of a formalism would be. Of course, one can preserve the watertight character of the static conception by regarding the analysis of equilibrium as a *purely logical* exercise in establishing the consistency of a number of assumptions or postulates, having no explanatory, predictive or even interpretative value. Silence is, similarly, the most effective way of avoiding making false statements.[2]

---

[1] It has been disputed that the static conception can even accommodate change. Joan Robinson has argued that one cannot approach an equilibrium: to be in it one must always have been in it. See Robinson, [1971*b*], p. 53: '. . . it is impossible for a system to *get into* a position of equilibrium, for the very nature of equilibrium is that the system is already in it, and has been in it for a certain length of past time'. (Emphasis in original.)

[2] Recent developments in general equilibrium theory have resulted in increasingly nebulous characterisations of the equilibrium state. For example, a concept of equilibrium is postulated which is consistent with the failure of markets to clear, provided they clear in some suitable stochastic sense, and consistent with the existence of learning

The issue of the preservation or violation of the strict logic of a formalism links up naturally with the next theme: the need for the expression of imprecise ideas in economics. This theme can be divided into two separate issues: first, the establishment of the thesis that, in economics, we are forced to deal with concepts characterised by an irreducible degree of imprecision; second, the raising of the question as to what medium is adequate to the proper expression of such ideas. To take the first issue, Shackle writes:

> A theory, to be useful, must be capable of being seen as exact, simple and essentially unchanging, and to be not misleading it must, in economics, be the vessel of cloudy, fluid and elusive conceptions.[1]

Further:

> Economics is the supremely ingenious device for eliciting scalar quantity from vast heterogeneous assemblies of incommensurable things. But this trick only serves certain purposes. It submerges detail, not abolishes it.[2]
>
> It [economics] is the art of heroic simplification, the art of the Gordian knot, carefully tied up in advance, however, by the very man who is going to cut it. But the economist ought to bear in mind that his conjuring tricks have their element of illusion, which he must not allow to deceive himself, despite the strong temptation.[3]

It is important to be quite clear what Shackle is here advocating: it is not a licence for muddled or slipshod or woolly thought. His urging of the need for the expression of imprecise ideas does not concern formulations which are a 'failed attempt to be precise'[4] but with deliberate and careful formulations which, as it were, mirror the inherent imprecision of the subject matter. If this account is sound, it leads to the seemingly paradoxical (but in fact straightforward) idea that carefully imprecise concepts can give a more *accurate* expression of the economic world than precise ones.[5] On these grounds, the kind of precision

---

on the part of individual traders, provided this learning is part of an equilibrium mode of adaptation to the economic circumstances. Such developments clearly have the effect of greatly weakening the force of the equilibrium concept. Although a proper consideration of these developments would make the discussion far more tangled, it would not actually undermine any of the conclusions: for despite any degree of weakening of the concept of equilibrium, the problems of pre-reconciliation in getting into the equilibrium state remain; and the formalisation of the general equilibrium system will still involve an *internal* logic which must be violated in order to put the formal scheme to any explanatory, interpretative or predictive use. For the investigation of the aggregate *consistency*, within a formal scheme, of the individual patterns of contingent intended responses to a set of circumstances, cannot consider anything other than the consistency of these patterns without violating the internal logic of the scheme. And this is evidently so no matter how refined or sophisticated the characterisation of the individual patterns may be, and no matter how guarded or circumscribed the notion of 'consistency' itself may become. (For a brief survey of recent developments in the conceptual basis of general equilibrium theory, see Hahn [1973].)

[1] Shackle [1972], p. 72. [2] *Ibid.*, p. 39. [3] *Ibid.* p. 10.

[4] Correspondence with the author, July 5, 1973.

[5] See Coddington [1969] where the problem of delimiting an economic concept in practice is compared with the problem of building a brick wall round a patch of fog. *Cf.* Shackle's '. . . sometimes its [economics] task is to find fixed relations amongst the wisps and eddies of a flowing mist', p. 177.

aimed at by the axiomatisers can be seen to be quite artificial in that to increase the precision of formalisms in no way contributes to a clarification of the mode of correspondence between the formalism and the economic world it is supposed to represent; and it is on the robustness of this correspondence that understanding (as opposed to the manipulation of symbols) ultimately rests. In fact the frequently-expressed view that the axiomatic method somehow puts an end to all ambiguity in our formulations rests on a confusion between the syntactic and semantic aspects of a formal scheme. The axiomatic method ensures that the formalism is severely disciplined in its syntactic aspect; but 'ambiguity' concerns its semantic properties, the clarity of the particular interpretation we are unable to give to the whole exercise.

As the virtues of quantity and measurement hardly lack spokesmen and recognition, Shackle concentrates on their dangers. He writes:

> Quantity obscures form. To be always considering ratios is to be in danger of being blind to the subtle and intricate *jigsaw-puzzle problems* with which engineering, marketing, labour relations and all the real and concrete detail of business confront the businessman. Midas could touch nothing without turning it to gold, and the economist likewise turns everything to sets of proportions. This relegation of form, structure and detail in favour of addable scalar quantities is his characteristic method of simplification, it is the essence of the market method which itself is the sole excuse for separating economic considerations from others.[1]

Further:

> Economics is the science which, of all scholarly disciplines, most recklessly oversteps the gulf between the humanities and the physical sciences. Making a virtue of the imprecision which its double face imposes on it, it has developed the technique of index-numbers into an engrossing and subtle branch of practical mathematics. Seizing on the peculiar character of markets, which draw men of widely different tastes into apparent unanimity concerning the proper mutual exchange value of goods, it has declared these public values to have the force and meaning of such attributes as length and mass,[2] and to be able to support a vast structure of aggregative calculations. Economics has virtually turned imprecision itself into a science: economics, the science of the quantification of the unquantifiable and the aggregation of the incompatible. It has followed this road at so violent a gallop, that much which is of significance and influence has been trampled on, *much territory has been claimed which cannot be held.*[3]

Again:

> By its insistence on measuring the unmeasurable, economics bludgeons

---

[1] P, 39, emphasis in original.
Cf. 'A horse and a man
Is more than one
And yet not many'
The Taming of the Shrew, Act 3, Scene 2, lines 82–4.

[2] See also Shackle [1972], p. 46.

[3] *Ibid.*, p. 360, emphasis added.

> the face of reality, the detail and vitality of human concerns are flattened to an unrecognisable tedium.[1]

Rather than taking up a discussion which cannot possibly be adequately tackled here, I would like to add a further three brief quotations to those from Shackle, and then leave the whole issue in mid-air. The first is from Aristotle:

> Our discussion will be adequate if it has as much clearance as the subject-matter admits of, for precision is not to be sought for alike in all discussion ... for it is the mark of an educated man to look for precision in each class of things just so far as the nature of the subject admits . . .[2]

The second is from Gauss: 'Lack of mathematical culture is revealed nowhere so conspicuously as in meaningless precision in numerical computations.[3]

The third is a remark of Frank Knight, to the effect that the practical meaning, for the social scientist, of Lord Kelvin's[4] dictum 'If you cannot measure, your knowledge is meager and unsatisfactory'[5] is 'If you cannot measure, measure anyhow'.[6]

I realise that these fragments do not constitute an argument, nor even the embryo of an argument. Rather they point, at most, to the absence of an argument: one concerning the status and significance of quantitative economics. The rôle which such an argument should have played has been played instead by its understudy, enthusiasm.[7]

I turn now to the theme for which Shackle is perhaps best known: the criticism of the concept of probability as a device for coping with the kind of problems—the expression of uncertainty and imprecision—which Shackle himself has been raising. Since this aspect of Shackle's work has been expounded at length in his previous work, and is reasonably well-known (if not well-appreciated), I will touch on it only briefly here, making the point first, that this now appears as an integral part of a critique of a much wider body of ideas than the previous expositions, and second, that in this statement of his position, Shackle surpasses all his previous ones in force and coherence.

Without in any way trying to accomplish the impossible task of summarising or condensing Shackle's argument here, it might be possible to gain some insight into (or sympathy with) Shackle's viewpoint by considering the curiosity of the grammatical form of probability statements. The statement 'The probability that this experiment of type A will result in an event of type B is p' appears superficially to be a statement about a *particular case of experiment* A (the case just about to take place). In fact, under analysis, the meaning of the statement turns out to be quite different from anything suggested by its grammatical form, for it turns out that the statement is not about a particular case of experiment A at all, but about a class of such cases (exactly what class of cases is problematic, and is something any interpretation of probability has to try to clear up).[1]

---

[1] *Ibid.*, p. 111. See also especially pp. 46, 49, 104, 177.

[2] Aristotle *Nicomancean Ethics*, Ch. 3, 1094*b*, 12–28. Translated by W. D. Ross.

[3] Quoted in Morgenstern [1963], p. 99.

[4] In fact there is no record of Kelvin ever having expressed precisely this sentiment, although it is within the spirit of his views.

[5] Inscribed on the façade of the Social Science Research Building at the University of Chicago.

[6] Quoted in Kuhn, [1961], p. 164.

[7] This enthusiasm seems largely to have been directed towards imitating what was taken to be the distinctive feature of physical science. It is, however, arguable that what has been imitated is a surface feature (see, *e.g.* Wiener, [1964]). It is perhaps worth mentioning that when I was an undergraduate student of physics, the first thing I learned was how to make rough and ready calculations of the orders of magnitude of the uncertainty in the results of operations and calculations, and how to express these ideas by such devices as dropped digits (as a half-way house to rounding off the digit). At no stage was I taught, or required to learn, the theory of statistics, nor the probability calculus. I relate this not for the autobiographical detail, but for its illustration of the contrast between the physicist's attitude to the craft of handling quantities, and the training which economists get in what is nowadays referred to as 'numeracy'.

Of course, anyone who knows anything about probability theory would realise that probability statements have a misleading grammatical form, even if he might not put it in those words. Shackle's objection to the deployment of probability theory can be seen as directed against those who, perhaps despite a deeper awareness of what is involved, have allowed themselves to venture onto the slippery slope of considering probability statements in terms of their (grammatical) face value. It is a slippery slope because probability ideas are desperately convenient for skating over or evading the kind of deep (and perhaps ultimately insoluble)[2] problems which Shackle has taken it upon himself to confront. Hence his characterisation of the word probability having been 'seized on as an incantation'.[3] For probability is not about uncertainty or imprecision or anything of that kind; insofar as it can be made operational, it is about the ratios of the numbers of (relevantly similar) members in different classes. Far from being about uncertainty, it is a piece of information; far from being about epistemics, it turns out to be 'a branch of arithmetic'.[4]

It is in the interpretation of Keynes's *General Theory* that we arrive at the *motif* which I have used in the title of this review. Here I cannot do better than quote directly from Shackle:

> The *General Theory* is, throughout, in two minds. It turns instinctively towards stable functions, uninterrupted movement along curves, underemployment 'equilibrium', secular stagnation, step-by-step declension. . . . Yet the message spelled out by all this creaking semaphore is that intended (designed, *ex ante*) investment is a law to itself, dependent (if at all) on too elusive a skein of subtle influences, too eagerly clutching at the straws of suggestion whirled along by 'the news', to be ever captured in any intelligible, let alone determinable, equation. It is not really the shape of the curves, but their broad bodily shifts and deformations, that contains the meaning of the argument'.[5]

Again, referring to the *General Theory*:

---

[1] Of course, Keynes's interpretation of probability as a relationship between statements, in general falling short of entailment, is a different approach altogether, but since it gives no practical guidance in the assigning of probability numbers, it need not concern us here. See Keynes, (1951).

[2] But if problems are ultimately insoluble it may be wiser to admit this than adopt a mode of expression which suggests that they have been solved.

[3] See above, p. 5 and Shackle [1972], p. 16.

[4] See *Ibid.*, p. 408.

[5] *Ibid.*, p. 218.

> Situations are portrayed by curves (or their equivalent) in two dimensions, connecting at most three variables . . . But the meaning of these situations is that of momentary, ephemeral glimpses at selected and rare points of a mainly un-adjusted, groping and speculative process, involving vast numbers of variables subject in many cases to an inherent restlessness and precariousness. The jump from one aspect to the other of his two-faced presentation consists formally in the liability of the curves to abrupt unexplained shifts. We do not ask for explanations of these shifts except on the most *ad hoc* basis, appealing to 'special factors'. The type of theory we are describing renounces, of its nature, any claim to offer a general, permanent frame of explanation of the shifts . . . But if theory pretends only to give an account of particular, peculiar and special moments . . . and repudiates any hope of connecting them by any intelligible permanent mechanism allowing prognosis, then the theory ought explicitly to be a classificatory one, putting situations in this box or that according to what *can happen* as a sequel to it. Theories which tell us what *will* happen are claiming too much . . .[1]

The trouble was that Keynes's 'creaking semaphore' was converted into a well-oiled and more elaborate variety by his systematisers and expositors (Hicks, Samuelson, Hanson, *etc.*), with the result that the message spelled out was, if Shackle is correct, made even more difficult to interpret. The 'two minds' of the *General Theory*, poised precariously between intellectual nihilism and a grasping for system and order, became submerged in a shallow and mechanistic synthesis.[2] It is this 'sophisticated semaphore', I believe, which was responsible for the post-Keynesian enthusiasm for fiscal management of the economy, an enthusiasm which, in retrospect, now appears something of an embarrassment.

A typical response to the collapse of mechanistic Keynesianism is to find a methodological scapegoat. The failure of theorising in its (policy) relevance and realism is castigated[3]; the inadequacy of available facts is blamed[4]; the villain is said to be the concept of equilibrium.[5] All these things (and others) are said to hold back and immobilise applied economics in its quest for effective economic policies. But there is something odd about such protestations coming from applied economists with a professional interest in policy questions. The nagging question arises: why do they raise these grumbles now rather than at any other time during the last twenty years or so, when the grounds for the complaints would have had just as much force? And most obvious of all: what right have such people to *complain* about the failure to produce what it is, after all, their job to produce?

---

[1] *Ibid.*, pp. 72–3, emphasis in original.

[2] It may be that 'hydraulic' (or, properly, 'hydraulistic') would be a better designation than 'mechanistic', on the grounds that the class of models all postulate determinate causal relationships between overall *flows*. What I refer to as 'mechanistic Keynesianism' corresponds reasonably well to what Leijonhufvud calls the 'income-expenditure model' (although this is not so much a model as a class of closely-related models). See Leijonhufvud, *op. cit.*

[3] See *e.g.* Worswick [1972]. This failure is said to arise from 'an imbalance among the specialisms which go to make up economic science' (*Ibid.*, p. 83).

[4] See *e.g.* Leontief, [1971].

[5] See *e.g.* Kaldor, [1972].

As against these calls for 'realism', 'relevance', 'facts' and so on, Shackle, in reflecting on the nature of his subject, offers the following piece of wisdom on the proper aspiration of the economics profession:

> A discipline, a region of the world of thought, should seek to *know itself.* Like the individual human being, it has received from its origins a stamp of character, a native mode of response to the situations confronting it. Right responses, 'responsibility' will require of the profession as of the individual an insight into the powers and the defects of the tool which history has bequeathed to it.[1]

ALAN CODDINGTON
Queen Mary College
University of London
and
University of Manchester

REFERENCES

Coddington, A. [1969]: 'Are Statistics Vital?', *The Listener*, **82**, pp. 822–3.

Coddington, A. [1972]: 'Positive Economics', *Canadian Journal of Economics*, **5**, pp. 1–15.

Hahn, F. H. [1973]: *On the Notion of Equilibrium in Economics.*

Kaldor, N. [1972]: 'The Irrelevance of Equilibrium Economics', *Economic Journal*, **82**, pp. 1237–55.

Keynes, J. M. [1921]: *A Treatise on Probability.*

Kuhn, T. S., [1961]: 'The Function of Measurement in Modern Physical Science', *ISIS*, **52**, pp. 161–93.

Leijonhufvud, A. [1968]: *On Keynesian Economics and the Economics of Keynes.*

Leontief, W. W. [1971]: 'Theoretical Assumptions and Non-observed Facts', *American Economic Review*, **61**, pp. 1–7.

Marshall, A. [1920]: *Principles of Economics.* Eighth edition.

Morgenstern, O. [1963]: *On the Accuracy of Economic Observations.* Second edition.

Robinson, J. [1971*a*]: *Economic Heresies.*

Robinson, J. [1971*b*]: 'The Production Function and the Theory of Capital', in G. C. Harcourt and N. F. Laing (*eds.*): *Capital and Growth*, pp. 47–64.

Shackle, G. L. S. [1972]: *Epistemics and Economics: A Critique of Economic Doctrines.*

Wiener, N. [1964]: *God and Golem Inc.*

Worswick, G. D. N. [1972]: 'Is Progress in Economic Science Possible?', *Economic Journal*, **82**, pp. 73–86.

---

[1] Shackle, [1972], p. 24, emphasis in original.

Brit. J. Phil. Sci. **26** (1975), 165–182 *Printed in Great Britain*

Reviews

FRAENKEL, A. A., BAR HILLEL, Y. and LEVY, A. (with the collaboration of D. van Dalen) [1973]: *Foundations of Set Theory*. Second, revised edition. Amsterdam: North-Holland. £10.00. Pp. x+404.

Of all the many branches of mathematics, set theory has the closest links with logic and philosophy. Accordingly, an author who attempts to give an adequate discussion of the *foundations* of set theory is confronted with a formidable task, for he must not only describe the purely *technical* aspects of the subject, but must also elucidate the role that logic and philosophy have played in its development. In the present book—a revised and expanded edition of the well-known work first published in 1958—the authors have taken on this task and, as in the earlier edition, produced a clearly written and remarkably comprehensive account of the development and present status of set theory. In this edition major revisions have been made in Chapter II on the axiomatic foundations of set theory, Chapter IV on the intuitionistic approach, and in the sections on Quine's systems and the limitative theorems of Gödel, Tarski, *et al.* There is also an excellent bibliography.

The book begins with a critical and historical discussion of the set-theoretic and logical antinomies which appeared at the end of the nineteenth century, and the ensuing 'crisis' in the foundations of mathematics. The various methods of resolving this crisis are classified by the authors into three main groups, *viz.*, the *axiomatic*, the *logicistic*, and the *intuitionistic* methods. Whereas the first method, and, to a certain extent, the second, are basically designed to salvage as much of 'Cantor's paradise' as possible, the third method involves a rejection of the greater part of its concepts and methods.

Chapter II is concerned with the first of these methods of resolving the foundational crisis caused by the discovery of the antinomies: the *axiomatic* method. The authors characterise this method as follows:

> 'the axiomatic attitude towards set theory . . . differs from the logicistic and intuitionistic attitudes not in that the latter are less strict in demanding a rigorous development of set theory, but in that it believes in the soundness of logic as used in mathematics through the ages and views the logical—*i.e.* set-theoretic—antinomies not as a failure of logic but only as a failure of Cantor's basic assumption about sets as expressed in his "definition" of a set. Therefore the cure advocated by the axiomatic attitude is to formulate new axioms in a way which will, at least apparently, avoid the occurrence of antinomies.'

The chapter begins with a discussion of Zermelo's set theory, as modified by Fraenkel and Skolem. With scant justice to the latter, this modified system is called Zermelo–Fraenkel set theory (*ZF*); it is the system of set theory most familiar to mathematicians. The authors show how the axioms of *ZF* (with the exception of the axioms of extensionality and foundation) are all particular instances of the principle whose unrestricted use was the original source of the set-theoretical antinomies, namely, the *comprehension principle* C: for any set theoretic condition $\phi(x)$, there is a set which contains exactly those elements x which satisfy this condition. The selection of these particular instances of C as axioms for ZF is guided by what Russell called the *limitation of size doctrine*, namely, to admit as axioms only those instances of C which assert the existence of sets which are not too 'big' in comparison with sets already obtained. Thus the effect of the limitation of size doctrine is to exclude those instances of C which give rise to large inconsistent 'sets', *e.g.* the 'set' of all ordinals or the 'set' of all sets not members of themselves.

The authors properly devote much attention to the *axiom of choice*, which of course asserts the existence, for each family A of non-empty sets, of a function on A which 'chooses' an element of each set in A. (This axiom was given its first explicit formulation by Zermelo in 1904 in connection with his proof of the well-ordering theorem.) Although—as the authors point out—a sufficiently liberal interpretation of Cantor's famous 'definition' of set (i.e. 'a collection into a whole of definite distinct objects of our intuition or of our thought') can arguably be used to justify the axiom, it is clear that, owing to the arbitrary nature of the 'choice' of an element from each set, the axiom is not an instance of the comprehension principle. In fact the axiom is of a purely existential character, *i.e.* it asserts the *existence* of a certain object (a choice function) but does not, in general, provide or even indicate a means of *constructing* it. This latter fact led the intuitionists—who of course insist that one can assert the existence of a mathematical object only when one has the wherewithal to construct it—to reject the general form of the axiom. The authors make the shrewd observation that, with the exception of the intuitionists, both the defenders and the opponents of the axiom were 'far more strongly influenced by emotional or practical reasons than by arguments of principle'. For those accepting and using the axiom, its chief justification is its 'indispensibility for proving important theorems of analysis and set theory'. (Although it should be pointed out that some mathematicians regard it as self-evidently true!) Interestingly enough, some of its early opponents also *rejected* it largely because of its consequences, in particular, the fact that it implies on the one hand that the continuum can be well-ordered (*i.e.* its cardinality is an aleph) but sheds no light on the position of this cardinality in the transfinite series of alephs (the continuum problem). The authors acutely observe, however, that 'this scepticism [with regard to the axiom] is based upon not appreciating the existential character of the well-ordering theorem which *a priori* makes its usefulness for the decision of "constructive" questions like the continuum problem highly doubtful'. They state that, in fact '. . . one has arrived at the axiom of choice just as at other mathematical principles, *viz.*, by *a posteriori* examining and analysing concepts, methods, and proofs actually found in mathematics whose original development in an intuitive manner rests on psychological rather than logical foundations'. They tentatively adopt a *pragmatic* attitude towards the axiom: '. . . namely, not to ask whether the axiom of choice is 'true', unavoidable, or admissible, but [to determine] what additional parts of mathematics can be obtained by the admission of the axiom'. Thus the authors believe that *mathematical practice*, rather than *metaphysics*, determines the acceptability of the axiom of choice.

The authors take a similarly pragmatic point of view with regard to the (generalised) ***continuum hypothesis***, which is discussed later in the chapter under the general heading 'Questions Unanswered by Axioms'. (In the presence of the axiom of choice, the generalised continuum hypothesis asserts that the cardinality of the power set of an infinite set A is the immediate successor of the cardinality of A; the continuum hypothesis is the special case in which A is the set of natural numbers.) They argue that '[Gödel's] hope for a more profound understanding of the notion of set resulting in some new axioms which settle the continuum problem has not been fulfilled', and that in any case 'one can hardly believe that many mathematicians will be inclined to accept some new axiom solely, or mainly, on the basis of a metaphysical belief in its power to reveal the true nature of the concept of set'. On the contrary, the authors feel that the generalised continuum hypothesis itself is well on its way towards being accepted as an axiom of set theory because of its usefulness in proving mathematical theorems and the fact that none of its alternatives are known to produce any mathematically interesting results. This may be true, but nonetheless many logicians regard the generalised continuum hypothesis as having much too *ad hoc* a character to make it acceptable as an axiom of set theory, and still hope to find a solution to the continuum problem along Gödel's lines.

After a discussion of large cardinals and strong axioms of infinity, Chapter II concludes with an exposition of these systems of set theory in which ***proper classes*** (***i.e.*** very large sets) are countenanced. These include the well-known systems of von Neumann and Bernays, and Quine and Morse. The choice of axioms for these systems are also guided by the limitation of size doctrine, and accordingly they are all, in a certain sense, natural extensions of ZF. Also mentioned is the less familiar but extremely interesting system of Ackermann, which differs from the others in that its main set formation principle is not derived from the limitation of size doctrine (although it is an instance of the comprehension principle). In Ackermann's system the domain of discourse consists of ***classes*** and there is, in addition to the relations of equality and membership, a unary predicate M, where $M(X)$ is understood to signify that the class X is a ***set***. Now Ackermann's key set formation principle asserts that, if $\phi(X)$ is any formula of the system which (1) does not involve the predicate M and is such that (2) the only classes which satisfy $\phi(X)$ are ***sets***, then the class of all sets which satisfy $\phi(X)$ is itself a set. Thus, if we think of sets as the extensions of 'well-defined' properties, Ackermann regards a property of sets as well-defined provided it fails to make explicit reference to the concept of set. However, since one of the other axioms implies that the collection of all sets is a class, it is clear that a formula containing bound class variables, even if it does not refer ***explicitly*** to the set concept, will nonetheless refer ***implicitly*** to this concept. Why a property involving an implicit reference to the set concept should be any more 'well-defined' than one involving an explicit reference was not satisfactorily explained by Ackermann. Despite these methodological reservations, however, Ackermann's system remains a strikingly original contribution to axiomatic set theory. Moreover, its strength and soundness have been amply demonstrated in view of the fact that it is now known to be ***equivalent*** to ZF. This remarkable result was established by Lévy and Reinhardt.

The systems discussed in Chapter II have the common feature that they are all formulable in the first-order predicate calculus. This means that the domain of discourse of each of them contains exactly one basic kind of object: *sets* in the case of *ZF*-style systems, and *classes* in the case of Von Neumann–Bernays style systems. Moreover, *any* expression of the form $x \in y$ or $x = y$, where x and y are (names for) variables is admitted as well-formed. In these systems the antinomies are avoided by restricting the *axioms*, while the syntax and *logical rules* of the first order predicate calculus are retained. *Chapter III* is mainly devoted to those systems—the *type-theoretical* systems—in which the antinomies are circumvented by modifying the syntax and possibly the logical rules of the first-order predicate calculus. (These systems, according to the authors' terminology, also fall under the heading of *logicistic* systems because historically they arose from an attempt to reduce set theory to *logic*.) The most familiar system of this kind is of course that presented by Russell and Whitehead in their celebrated work *Principia Mathematica* (1910–13). The underlying language of this system, which we shall refer to as *PM*, includes a countable hierarchy of *types* of variables: $x^1, y^1, z^1, \ldots; x^2, y^2, z^2, \ldots; \ldots; x^n, y^n, z^n, \ldots; \ldots.$ The variables x^n, y^n, $z^n, \ldots$ are understood to range over sets of *type n*. Only expressions of the form $x^i \in y^j$, where $j = i+1$, or $x^i = y^i$ are admitted as atomic formulas. That is, in *PM* one can only assert that a given set is or is not a member of (equal to) another *provided* the first set is of type one lower than (has the same type as) the second. The axiom scheme of comprehension now assumes the form

$$\exists y^{i+1} \forall x^i [x^i \in y^{i+1} \leftrightarrow \phi(x^i)],$$

where ϕ is any formula. In other words, given a property of sets of type i, there is a set of type $i+1$ consisting precisely of all those sets of type i which have the property. It is easy to see that the usual antinomies are circumvented in this system and it was shown by Russell and Whitehead that, if one adjoins a suitable axiom of infinity, one can develop all of classical mathematics within it. Its main drawback is the unnatural duplication of entities (*e.g.* the empty set, cardinal numbers, *etc.*) at each type level, which, among other things, makes it a rather cumbersome framework in which to develop classical mathematics. On the other hand, the natural way it avoids the antinomies gives rise to a considerable degree of confidence in its consistency. The authors discuss this and several other type-theoretical systems, including Quine's New Foundations (*NF*) and Mathematical Logic (*ML*), the Ramified Class Calculus (*RC*) and Wang's system Σ. The main trouble with all these systems is that, in general, it is difficult (and sometimes impossible) to develop important areas of classical mathematics within them. For example, Cantor's theorem that the cardinality of any set is strictly smaller than that of its power set fails in both *NF* and Σ!

Chapter III concludes with a discussion of some of the attempts that have made to build set theory within a framework obtained by changing the underlying *logic*, for example, Chwistek's and Fitch's systems. The most important attempt of this kind, however, is that mounted by Brouwer and the intuitionistic school. The intuitionistic approach to mathematics forms the subject-matter of *Chapter IV*.

The basic principles of intuitionism may be roughly stated as follows: (1) a mathematical object exists if and only if it can be *constructed* (in some definite way), the construction constituting the proof of existence; and (2) an infinite totality must never be treated as if it were a completed whole, but must rather be regarded as continually 'growing'. It is well known that these two principles, if taken seriously, have far-reaching implications for logic and mathematics. For example, (1) entails the rejection of the principle of the excluded middle for infinite totalities, and (2) necessitates not only a radical revision of the classical concept of the continuum, but the outright rejection of uncountable infinities. Accordingly, intuitionism regards as *meaningless* that part of Cantor's theory which deals with uncountable sets, and deals only with those parts of set theory (arithmetic and analysis) which can be reconstituted along constructivistic lines. (In particular, the set-theoretic antinomies are avoided by denying the relevant mathematical concepts any meaning!) The authors give a thorough discussion of how this reconstruction has been carried out, and the extent to which the classical theory can be saved. Moreover, there is a very clear and (to the reviewer, who is no intuitionist!) enlightening discussion of the general development of intuitionistic ideas and their impact on mathematics. Also included is a concise and lucid description of Brouwer's (notoriously obscure) philosophical views. All in all, this chapter forms a good introduction to intuitionism, which may be profitably studied by any mathematician interested in acquainting himself with its foundations.

The concluding chapter of the book is chiefly designed to give the reader enough technical background in mathematical logic to grasp the major results in the metamathematics of set theory. The authors begin the chapter with a discussion of Hilbert's programme and his development of the formal axiomatic method. Later on we find an account of the Löwenheim–Skolem theorem, Tarski's theorem on the undefinability of truth, Gödel's incompleteness theorem, and Church's undecidability theorem. The authors discuss in detail how these and other metamathematical results have affected the development of axiomatic set theory, and, more generally, how they establish the essential limitations of the axiomatic method. The authors round off their technical discussion with a summary of the relative proof-theoretic strengths of many of the systems of axiomatic set theory discussed in earlier chapters.

The final pages of the text are devoted to the philosophical issues raised by problems in the foundations of set theory. After pointing out that the major problem in the philosophy of set theory is that of the *ontological* status of sets, they go on to discuss the four major positions which have been held with regard to this question: *Platonism*, which regards sets as existing objectively and independently of human activity, *nominalism*, which regards the universe as composed of individuals and sets as a mere *facon de parler*, *conceptualism*, which regards sets as produced by more or less definite mental constructions, and *anti-ontologism*, which regards the whole ontological issue as irrelevant to the foundations of set theory. (Of course, these positions are not all mutually exclusive!) The authors themselves appear to favour the fourth of these positions, in particular, they believe that mathematical results are *neutral* with respect to ontological questions. Indeed, they assert that: 'should one want . . . to conclude that all ontological views on mathematics, since irrefutable, are thereby also irrelevant for mathematics, though not necessarily for mathematicians, we see no good reasons against such a conclusion'. Similarly, with respect to the related question of the ***uniqueness*** of set theory, they remark that, although it is easy to understand the urge for such a belief, 'so far we do not see any reason compelling us to believe that there will be a unique solution to the foundational problems of set theory which will induce all mathematicians to accept one such theory as ***the*** Set Theory'. They sound the warning that

> . . . there is often only one step from the belief in the existence of an objective criterion that would uniquely determine the issue between competing theories and the belief that one has found this criterion and is therefore entitled to disqualify all those theories, except possibly one, in the name of some earthly or heavenly reality.

Finally, however, they reassure one by observing that

> although the foundations of set theory are still somewhat shaky, . . . mathematicians continue to apply successfully its concepts, methods, and results in most branches of analysis and geometry as well as in some parts of arithmetic and algebra, confident that future foundational research will converge towards a vindication of set theory to an extent that will be identical with, or at least close to, its classical one.

Thus, although the ***philosophy*** of set theory may still be in something of a muddle, as a ***branch of mathematics*** it is alive and well.

To sum up, this book is a masterly survey of its field. It is lucid and concise on a technical level, it covers the historical ground admirably, and it gives a sensible account of the various philosophical positions associated with the development of the subject. This work is essential reading for any mathematician or philosopher—professional or prospective—who has an interest in the foundations of mathematics.

J. L. BELL
London School of Economics

WANG, HAO [1974]: *From Mathematics to Philosophy*. London: Routledge and Kegan Paul in the series International Library of Philosophy and Scientific Method. £7.75. Pp. xiv+428.

Wang's book is composed of a collection of previously published papers, some of which have been revised and extended, together with four chapters unpublished elsewhere. Most of the chapters deal with topics related to mathematical logic and the foundations of mathematics: Chapters I, 'Mathematical Logic and the Philosophy of Mathematics'; II, 'Characterisation of general mathematical concepts'; III, 'Russell's logic and some general issues'; IV, 'Logical truth'; V, 'Metalogic'; VI, 'The Concept of Set'; VII, 'Theory and practice in Mathematics'; VIII, 'Necessity, analyticity, and apriority'; IX, 'Mathematics and Computers'; X, 'Minds and Machines'. The Preface, the Introduction and Chapters XI, 'Notes on knowledge and life', and XII, 'Themes and approaches', are meant to provide the philosophical setting for the rest of the book. In addition, the Introduction contains long quotations from letters written by Gödel to Wang in 1967 and 1968 concerning Gödel's work in mathematical logic and set theory; and there are various places in which Wang states Gödel's (unpublished) views (especially on mechanical procedures and minds and machines) with Gödel's approval. The book has a fairly broad scope, and as Wang states, although the chapters are loosely interconnected, they can all be read more or less independently. However he also states that the book as a whole is meant to 'illustrate a general position on the philosophy of knowledge', a position which he attempts to set out in the philosophical chapters.

This 'general position' however seems rather vague and confusing, not to say naive. The basis of Wang's position is what he calls 'substantial factualism', the aim of which is to concentrate attention on 'what we know', *i.e.* on what he calls the 'brute facts' we possess, particularly in the physical sciences, logic and mathematics. Part of Wang's purpose, as he states, is to attack positivism, especially logical positivism in the philosophy of science and formalism and finitism in the philosophy of mathematics. He argues that these positions maintain a far too restrictive view of what is acceptable as knowledge. Wang may be right; but he does not make it clear whether he interprets knowledge as 'certain knowledge' or 'conjectual knowledge', or what. His argument is simply that, as he puts it, 'we know what we know'. Thus he completely ignores the age-old debate surrounding the epistemological problem: 'What justification do we have for saying that we know something?' In some parts of his book Wang implicitly assumes that the problem has been satisfactorily solved; while in other parts he explicitly assumes that it can be solved in terms of, for example, the 'rich experience from introspection' or 'cultivated intuition'.

In contrast to the vagueness and confusion of the philosophical sections, the book does have some interesting and extremely clear discussions of various problems in logic and the foundations of mathematics. The sections on axiomatics, consistency, finitism and set theory are particularly worth mentioning since they all reflect Wang's anti-formalism in a far more convincing way than the philosophical sections.

In the first place, Wang clearly points out the limitations of formalisation and the use of the axiomatic method in general. In his treatment of axiomatics (Chapter 1, Section 2) Wang expresses admiration for the work of the Bourbaki group. He stresses that the Bourbaki project is not simply an attempt to write out the whole of mathematics in one language, but an attempt to unify mathematics in a hierarchy of structures. The idea is that for a given theorem one tries to find the most general structure in which that theorem holds; in other words, to find the weakest possible conjunction of hypotheses which are necessary in the statement of the theorem. Yet, says Wang, this extremely abstract approach has obvious dangers. For one thing, it tends to divorce mathematics from what he calls 'mathematical reality' and particular theories appear to get lost, cut off from their problem backgrounds. Perhaps the biggest danger in this respect, although Wang strangely does not mention it, is the danger of teaching mathematics solely in its Bourbaki form. For instance, general topology or integration theory might appear completely meaningless to the student not familiar with the emergence of topology from real analysis, or with the problems in integration theory that led to Lebesgue's revision and extension of Riemann's theory.

Much the same point arises with the discussion of the axiom systems for the natural numbers and for sets. (Chapter II, section 1, and Chapter X, respectively). Wang, like Gödel, sees the axioms for set theory as an attempt to explicate the notion of 'set'. In the same way, as he points out, the Dedekind–Peano axioms for number theory were produced by Dedekind in an attempt to uniquely characterise the natural number series. Wang's point is simply this, that neither of these axiom systems should be considered in isolation from their intended models. Hilbert made this mistake with number theory, as Gödel subsequently pointed out. The point is important, and Wang makes it again in connection with intuitionism (Chapter 1, section 4). Much work has been done on trying to formalize intuitionistic reasoning, and on trying to write out intuitionistic mathematics in formal systems. Wang concedes that we can learn much about intuitionism through such attempts, especially by comparing these systems to the respective classical systems. However, as he stresses, these systems do not constitute intuitionist mathematics. As is well known, Brouwer himself was opposed to all such attempts at formalization.

Wang's anti-formalism is also reflected in his discussion of consistency (Chapter 1, section 3). For the formalists before Gödel's incompleteness results consistency was regarded as a necessary (and by some as a sufficient) condition for the adequacy of a formal system. After these results, formalists tended to regard the search for consistency proofs as fruitless. However, Wang stresses that consistency is perhaps necessary, definitely insufficient, and that a consistency proof can be enormously *fruitful*. For instance, the Gentzen consistency proof for Peano arithmetic enables us to characterise the power of the rule of mathematical induction in terms of ordinals. As Wang states, this proof reduces the reliability of mathematical induction in *all* its uses to the reliability of a *single* application of the rule of transfinite induction up to the countable ordinal ϵ_0. This reduction is extremely important. We can extract even more information from this proof. For example (a point not mentioned by Wang), it can be shown that given a standard method of coding, the coded statement expressing the rule of transfinite induction up to ϵ_0 cannot be proved in Peano arithmetic, whereas the coded statement expressing the rule of transfinite induction up to any ordinal less than ϵ_0 *can* be proved. This in some sense, must mean that arithmetic only just fails to prove its own consistency. None of these interesting results should have any meaning for the sincere formalist.

The strongest arguments against finitism in Wang's book are those used by Gödel in the letters quoted in the Introduction. Gödel attacks the sterility of the finitist position taken up by logicians of the Hilbert school (*i.e.* most logicians) in the 1920s. Gödel claims that he arrived at his fundamental completeness result because he had no qualms at all about using non-finitary reasoning, whereas the other logicians were prevented from achieving this result before him by their refusal to countenance non-finitary reasoning in metamathematics. The Hilbert programme set out to justify the use of any non-finitary reasoning in mathematics through finitary metamathematics, that is to say, through 'self-evidently correct' (finitary) proofs that the non-finitary elements cannot lead to inconsistency. Hence, non-finitary elements could not be allowed into metamathematics itself. Now, says Gödel, a version of the completeness theorem was stated in Skolem [1928], though Skolem himself was unable to find a satisfactory proof. However, the theorem as stated by Skolem can be proved easily by a simple application of a lemma Skolem had proved in his 1922 proof of Löwenheim's theorem. But Skolem did not use this lemma, and the reason must be, says Gödel, that the proof of the lemma used non-finitary reasoning. (It actually uses a weak version of the axiom of choice.) Moreover, Hilbert and Ackermann in their [1928] explicitly stated the standard completeness theorem (validity implies provability) as an open problem. The reason they too could not find a proof, says Gödel, is again due to their finitistic viewpoint. It was impossible for them to prove the theorem since, as Gödel points out, any proof of completeness for the predicate calculus must involve some non-finitary element. The completeness problem thus already contained a hint of a defeat for the formalists.

Gödel is undoubtedly correct as to the reasons why neither Skolem nor Hilbert and Ackermann could prove the completeness theorem; finitism turned out to be too restrictive. Yet Gödel's criticism seems unnecessarily harsh. In the first place, Skolem [1928] is very confused on the issue of completeness. Secondly, the Hilbert programme was a remarkably bold and wholly admirable project which logicians had no good reason to abandon until after 1931.

Gödel's letter constitutes perhaps the most interesting part of the book, and is worth considering at some length. After discussing the completeness problem he goes on: 'I may add that my objectivistic conception of mathematics and metamathematics in general, and of transfinite reasoning in particular, was fundamental also to my other work in logic'. Both his incompleteness results and his work on set theory are good examples of this 'conception' as Gödel points out. For instance, an essential step in the proof of his incompleteness theorem consists in mapping the metamathematics into the mathematical system itself. This procedure, which has since become an indispensible tool of mathematical logic, would have been unthinkable for any of the logicians working on the Hilbert programme, since for them the formal mathematical systems, meaningless in themselves, received meaning *only through* the metamathematics. For Gödel, clearly, it is the mathematical system itself which is meaningful. Moreover, says Gödel, his 1931 papers make essential use of the 'highly transfinite concept of "objective mathematical truth", as *opposed* to that of "demonstrability" '.

From these letters it is obvious that Gödel regards his papers [1938], [1939], [1940] as giving an *actual* (non-finitary) consistency proof for the generalised continuum hypothesis and the axiom of choice, and not merely a *relative* consistency proof. (Indeed the titles of his [1938], [1939], [1940] all begin 'The consistency of the axiom of choice . . . *etc.*') In other words Gödel obviously sees the constructible sets as furnishing an actual model of *ZF* plus *AC* and *GCH*. Certainly, by using a standard method of coding, Gödel's proof can be translated into a *finitary* relative consistency proof in simple arithmetic. But Gödel's point is clear; the non-finitary proof had to be discovered first. At least no mathematician can deny the *heuristic* power of Gödel's non-finitary reasoning.

Wang's chapter on set-theory is notable for two things: the excellent discussions first of the 'iterative concept' of set put forward by von Neumann in 1924, and taken up by Zermelo and Gödel; and secondly of Cantor's early work on set theory. In addition, Wang also puts forward two criteria (which are basically those advanced in Gödel [1947]) by which to judge proposed additions to the axioms of set theory. The two criteria are those of 'intrinsic necessity' and 'pragmatic success'. The appeal to 'intrinsic necessity', like the appeal to 'educated' or 'sophisticated' intuition stressed by Wang at various places in his book is not very helpful. No one can deny that 'intuition' or 'creative thought' is useful in *suggesting* extensions of set theory, or indeed in suggesting ways of testing extensions. However as a means of justification it can only lead to authoritarianism. The criterion of 'pragmatic success' seems more hopeful, yet as it stands, it is clearly not sufficient. According to Wang 'pragmatic success' might include the derivation of known lower-order consequences from a proposed new axiom, plus the derivation of unknown consequences. Yet there is nothing in this criterion to prevent the proposed new axiom from being completely *ad hoc.* There must surely be included some element of *unexpectedness*, that is, either some consequence must turn out to be unexpectedly confirmed (in a different, or related theory), or some unexpected consequences should be confirmed, preferably in some lower-order theory. Wang also fails to mention that the question of the introduction of new 'principles' or axioms into set theory nearly always arises in connection with some serious unsolved problem.

M. F. HALLETT
London School of Economics

REFERENCES

BENACERRAF, P. and PUTNAM, H. (*eds.*) [1964]: *Philosophy of Mathematics: Selected Readings*. Prentice-Hall.

GÖDEL, K. [1938]: 'The Consistency of the Axiom of Choice and of the Generalised Continuum Hypothesis', *Proceedings of the National Academy of Sciences*, **24**, pp. 556–7.

GÖDEL, K. [1939]: 'Consistency-Proof for the Generalised Continuum Hypothesis', *Proceedings of the National Academy of Sciences*, **25**, pp. 220–4.

GÖDEL, K. [1940]: *The Consistency of the Continuum Hypothesis*. Princeton University Press.

GÖDEL, K. [1947]: 'What is Cantor's Continuum Problem?', *The American Mathematical Monthly*, **54**, pp. 515–25. Reprinted as a revised version in Benacerraf, P. and Putnam, H. (*eds.*) [1964], pp. 258–73.

HEIJERNOORT, J. VAN (*ed.*) [1967]: *From Frege to Gödel*. Harvard University Press.

HILBERT, D. and ACKERMANN, W. [1928]: *Grundzüge der Theoretischen Logik*. Berlin: Springer.

SKOLEM, T. [1923]: 'Einige Bemerkungen zur Axiomatischen Begründung der Mengenlehre', *Matematikerkongressen i Helsinfors den 4–7 Juli 1922, Den Femte Skandinaviska Matematikerkongressen, Redogörelse*. Helsinki: Akademiska Bokhandlen. English translation in van Heijenoort (*ed.*) [1967], pp. 290–301.

SKOLEM, T. [1928]: 'Über die Mathematische Logik', *Norsk Matematisk Tidsskrift*, **10**, pp. 125–42. Reprinted in an English translation in J. van Heijenoort, (*ed.*) [1967], pp. 508–24. The editorial introduction to this translation also contains some remarks by Gödel on the completeness problem similar to those quoted *above*. See J. van Heijenoort, (*ed.*) [1967], p. 510, note i.

HOBBES, T. [1973]: *Critique du* De Mundo *de Thomas White*. Introduction, Texte Critique et Notes par Jean Jacquot et Harold Whitmore Jones. Paris: Librairie Philosophique de J. Vrin. Pp. 546.

An early work of Hobbes, written in Latin and hitherto unknown (except to a few of his contemporaries in France who saw it in manuscript), has recently been edited and printed from the MS preserved in the Bibliothèque Nationale in Paris. It consists of a systematic and detailed refutation of Thomas White's *De Mundo*, a scientifico-philosophical treatise published in Paris in 1642. Hobbes wrote his *Critique*, with the encouragement of his friend Mersenne, in the course of the following two years while he was an exile in Paris and in close contact with the intellectual circles there. In his Introduction M. Jacquot has provided an account of the essential background of the controversy, concentrating on historical, scientific, and literary information. My aim here is to offer a critical examination of Hobbes's leading ideas and arguments and to draw attention to those of special philosophical interest and relevance.

His first set of arguments proceeds as follows: he defines philosophy, in any of its applications (to the study of being in general, to physics, mathematics, ethics, politics, *etc.*) as the science of universal theorems, the truths of whose conclusions can be demonstrated of necessity by natural reason. The manner of presentation of this science must therefore proceed from the definitions of terms and the drawing out of what necessarily follows from such definitions. In this sense, then, for Hobbes philosophy is to be treated logically, in dissociation from both rhetoric and the art of eristic.

The first question which arises is whether the world is finite or infinite. It by no means follows from the meaning of these words that either attribute can be necessarily predicated of the universe in the way in which 'animality' is predicated of 'man'. Hobbes concludes that the question at issue, for lack of a necessary connection between the definitions under consideration, is incapable of solution. For in order to be infinite, the universe, a material body, or a number series must be capable of limitless extension. In the case of number it is possible to know from the meaning of the term that this extension can be achieved by means of the intellect and is therefore within man's power to accomplish. In the case of a body, however, though we can assume on rational grounds a perpetual addition of further parts to its dimension, no merely logical consideration can decide whether such indefinitely or infinitely increased body does in fact exist. The same applies in the case of the universe. From the definition of this word it is impossible to know anything about the sum total of all the material bodies that constitute the world, or whether it is possible to add any further parts to it. This sort of question can be answered only if one has knowledge of the divine will; to say that the world is finite is therefore an assertion dependent on religious faith and not one that can be determined by reasons advanced in science, philosophy, or logic.

Since Thomas White claims to be able to demonstrate, on equally firm grounds as a proof in mathematics, that the world is finite, Hobbes's endeavour is to show how this demonstration can be refuted. He does so on the following lines. White's argument—a syllogism—is that,

> If the world is infinite, some part of it is distant from us an infinite number of feet or some other such unit of measure;
>
> However no part of the world can be separated from us by an infinite number of feet;
>
> Therefore the world is not infinite.

Hobbes rejects the major premise of this ratiocination; for he rightly maintains that it is just as mistaken to infer that any part of an infinite world is itself infinite as it is to argue that a certain number is infinite simply because one can add one number to another *ad infinitum*. Hobbes made use of this part of his critique of White as of many others in subsequent scientific writings, particularly in his *De Corpore*. The fact that there was originally a specific opponent to his own mode of reasoning has left a few traces in his later work, for instance in *De Corpore*, Pt. II, ch. 7, sect 12,[1] where a personal (though nameless) reference has survived in the context of the argument I have mentioned.

There are of course reasonings in the light of which it can be shown that the world is infinite. Hobbes cites three, disputing the validity of White's attempts to refute them. The three arguments start from the assumption that what is (respectively) consistent, conceivable, or capable for God to create, is possible. One of White's counter-arguments is that the proposition 'the infinite is possible' conflicts with the well-known principle whereby singular things are subsumed under universals. Hobbes's nominalistic reply is that, since only words are universal which as such are the names of particulars so that a universal is something that is said or predicated either of other general words or of proper names, one can say that man is an animal and also that Socrates is an animal, but not that a man's hand is a man or that part of the infinite is infinite. Hence the original proposition about the infinite cannot be said to 'conflict' with the *dictum de omni* principle, since each is about something quite different, the one about the relation between part and whole, the other about the subsumption of the singular under the universal.

Another cosmological problem is whether there might be a plurality of worlds, or whether the world is one and single. Hobbes's critical examination of the second assumption—that defended by White—is based on the contention that spaces and positions of objects can be imagined, *i.e.* considered to appear to exist, outside the present world just as they may be imagined if the world, except for a single man, were supposed to be annihilated. He concludes that, from the concept of the universe even if this is intended as finite, it is no less possible to infer a plurality of worlds than it is to argue that the concept of sun or animal may imply a plurality of suns or animals respectively. For, as he puts it, from the fact that there is nothing it does not follow (as it does for White) that there is absolutely nothing; nor does it follow that where there is nothing (as in a merely imaginary space) there are neither parts nor divisions. Hence there are two possibilities, *scil.* that, if there are two worlds, either they are separate from each other with a certain space between them; or they are contiguous, in which case they are not (as White tries to make out) one world, or even a single compound or connected world, but still two different ones. But whether, if the universe is finite, there are *in fact* several worlds or only one, is (as before) not a subject-matter for philosophical reasoning but for religious faith.

A further topic for controversy is whether the outer heaven contains physical bodies similar to those existing on earth. Hobbes argues that, in order to account for variety it is necessary to accept as the principle of change something like the Aristotelian prime matter. Wood, for example, by itself is simply body, but it is also the matter of which a bench is made. Now prime matter, it could be maintained, is the most immutable and general element of body—though an abstraction really, not unlike the concept of 'objects' in Wittgenstein's *Tractatus*—for it lacks all determinate characteristics. Still, contrary to White, Hobbes believes that the number of possible alterations of prime matter and therefore of all material things is infinite; that consequently it is impossible to know how many changes there are in the world, nor whether among the heavenly bodies there are some which resemble certain objects on earth, or whether there are similarities between the moon as such and the earth. Being more traditionally-minded than White, Hobbes also dissociates himself from the latter's belief in the corruptibility of the heavenly bodies, indicated (it is alleged) both by the existence of sun spots which point to an eventual extinction of the sun's fiery substance, and also by the appearance of new stars.

---

[1] Translated in W. Molesworth (*ed.*): *The English Works of Thomas Hobbes*, I, p. 99, lines 18–24.

Hobbes attaches additional importance to the concept of matter because he considers it one of the principles of individuation. In philosophy there has always been the problem of what it is that constitutes the identity of an otherwise changing object or person during a given period of time. Descartes raised this question in connection with his experiment over the melting wax, and Hume dealt with it in his celebrated inquiry into personal identity. Hobbes's debate with White is of twofold interest. Firstly, in its essentials and in all its examples it has survived in a chapter of *De Corpore* (Pt. II, ch. 11), where there are also two veiled references to the person with whom Hobbes originally disputed this question and whose contribution to the discussion he rejects (*English Works*, vol. i, p. 136, lines 2 and 14). Secondly, it is more than probable that Hume derived from this chapter in *De Corpore* some of his own arguments about identity, *i.e.* his points about the same body, the unity of matter, the man grown old, the repaired ship, the river with its water always changing, and finally the difference between numerical and specific identity (*i.e.* that attributed to form rather than mere number). White's point is that identity depends on the mathematical unity of the aggregate of all the accidents of a thing together. Hobbes's reply is pertinent: firstly, he insists, no instance can be conceived of this criterion ever being satisfied, especially if (as White maintains) the variation of accidents may be infinitely great. Secondly, he considers that a person who does wrong and who is punished for it should not be (for reason of the constant changes within his body) numerically the same person—to the confusion of all civil rights and duties. Thirdly, it should be difficult to prove that the universe will be different in future simply because some of its particular features have been changed. For if the universe, on this account, does not remain one and the same, and yet neither it nor any of its parts can be held to perish as a result of such a minor alteration, then there must be two simultaneous universes in the same place—an absurdity.

Besides matter, a mechanist philosopher would inquire into the nature of motion. In the context of this investigation Hobbes takes up the concept of *conatus* or 'endeavour', which figures so prominently in all his subsequent writings on physics, psychology, ethics, and political philosophy. The reason why this key term occurs in Hobbes's critique at this point is that for White the world is finite, with a fixed outermost surface. Since there are turbulent movements within the universe, either its outer sphere must be exceptionally hard, or every motion and endeavour must be directed inwards. Following Copernicus and Galileo, White opts for the second alternative—a choice approved of by

Hobbes. It follows that all locomotion in the world must be either circular or directed inwards. In this connection Hobbes defines 'endeavour' as the actual beginning of motion or its infinitesimal, invisible, first part. For instance, a heavy body on the ground would endeavour to move downwards; for if what prevents this motion were removed the body would sink further down. Since the removal of the impediment itself is not an action on the part of the body, and since no body would begin to move without some action, it follows that the beginning of the descent of a heavy body is some actual endeavour on its part to descend.

Of equal if not greater significance for a proper understanding of motion and of nature as a whole is the question of whether movements are absolute or relative. In the case of absolute motion a body would leave one place for another; in that of relative motion there would be a mere change of position as between two bodies, so that either may be said to move—an interpretation vindicating talk in the Scriptures about the sun's motion. Whereas White is in favour of movement being relative, Hobbes is decidedly against it. His main point is that the sun, though stationary, may *appear* to move as a result of the change of relative positions between it and the earth. But to attribute motion to a body at rest on the grounds that it *seems* to our senses to move is not only inaccurate but logically improper, for the appearance of such a change depends—as in all optical matters (see *De Homine*, ch. II)—not only on the nature of the object perceived but also on the psychology of our own nature and on our perceptive faculties. Hence, in Hobbes's view, a real change only occurs if there is some truly new and absolute alteration in a thing. If John's son dies and John thereby ceases to be a father, John himself does not necessarily change—particularly if he is unaware of his son's death. Similarly, if the sun is seen to move by people on the earth or to have altered its position relative to that of the earth, this does not prove that the sun has in fact moved. The division of opinion here between White and Hobbes is not unlike that between Descartes's view of the reciprocity of movement and Newton's belief in the validity of statements about absolute space and motion.

The cause of motion is the next main topic of debate. Since this must give rise to strictly metaphysical questions, Hobbes introduces a few methodological warnings. Philosophers can establish true propositions if the meaning of the predicate terms includes that of the subject terms. Demonstrations of true propositions can be derived only if the inclusion in question is manifest from the definitions of the terms employed. Subject terms are of things *supposed* to exist, and conclusions have therefore only hypothetical, never categorical, force. However, whilst for demonstrations in connection with triangles no triangle need exist, in order to prove that something is part of reality experience is indispensable. It follows that in theological matters—those relating to the existence of God and the question of whether the world has been created or is eternal—faith, not natural reason or science has to be relied on; for a philosopher may no longer understand the true meaning of the terms used in this context. A position of fideism must therefore be adopted—a plea supported by the Aristotelian-Averroistic school of Padua whose disciple Hobbes was.

One of Hobbes's illustrations is the issue whether the motion of the universe has had an external first cause. The meaning of 'cause' is of course intelligible.

It signifies either the principle of efficient causation which resides in any thing acting upon another contiguous to it, or that of material causation—the correlative of the former—which lies in the thing acted upon. Final causes can be considered efficient ones, especially if a future purpose or goal becomes the present motivating force of an action. Formal causes are not strictly speaking 'causes' but conceptual implications following from our understanding of the properties of such a thing as a triangle. The question raised above, however, and answered in the affirmative by White is a metaphysical one and hence beyond human reason. This is particularly so, if there is a failure to recognise that 'universe' means the sum-total of all there is, so that it is absurd to conceive of it being moved 'from without' if there can be no 'without' and if, supposing the universe to be infinite, it cannot be 'moved'. Besides, it is perfectly intelligible to conceive of the world as *continuing* to be moved without any cause whatsoever. On the other hand, if what is 'outside' the universe (this being finite) is nothing, God as the creator of the world may equally be nothing—consequences too absurd to be considered by anyone who believes in the existence of God and who accepts the possibility of Creation.

Hobbes's own view is that, like the soul in man, something itself unmoved may originate motion in the universe. Like the soul (which can be described as the unitary and simultaneous series of all actions on the part of a man), an unmoved mover need not be explained in terms of immateriality or as being outside space. And whether motion in the world is or is not eternal, it can have had a beginning in the form of divine creation, with the divine will *determining* its exact time and place. This assumption naturally gives rise to the issue of determinism and that of the liberty of will in both God and man. Hobbes deals with these two important topics towards the end of his *Critique*.

He considers it perfectly intelligible to speak of God as having created the world of his own free will at a certain time and in a certain place, and with no antecedent conditions determining his choice. Nevertheless, he regards it as conceivable that there may have been moments of time and indeed an infinite duration of time preceding the act of creation. The argument (advanced by White and subsequently by Leibniz in his correspondence with Clarke) that there is no sufficient reason why God, when there were as yet no temporal determinations in the world, should have created it at one instant rather than another leaves Hobbes unimpressed: for him it is part of the nature of an eternal being (*i.e.* one who exists *a se*) to be free to choose, to act, and to cause everything in the universe at a determinate moment and in one place rather than another. Also, whereas men, when they want to act, will the action because it is good for them, in God's case an action is good because He decides to perform it. (Hobbes's approach here is anti-Platonic, nominalist, or Calvinist, not unlike Wittgenstein's: see F. Waismann, *Wittgenstein und der Wiener Kreis*, Oxford, 1967, p. 115.) Generally, the meaning of man's freedom is quite different from the sense given by Hobbes to that term in relation to God. In his view a person's will is never free but determined. What is free in the life of a human (or animal) agent is, first, the process of deliberation preceding and determining the will. This freedom of the *agent* means that, while considering his options and alternating (or oscillating) between an *endeavour* to do and an *endeavour* not to do, he has as yet not chosen how to act, and that he remains free as long as he is thus deliberating.

But secondly, an *action* (any action in fact, including inanimate motion) is free as long as it meets with no external impediment. Here, in the context of his *Critique* of White, Hobbes's future and complex doctrine of freedom and necessity is already present. The two kinds of liberty to which he refers, it must be admitted, seem to be incompatible, since on Hobbes's view a man can be said to be free in two quite different senses, one applying before and the other after he wills, and yet he is not free to will or free in willing. The incompatibility disappears only if one either maintains a hard and fast distinction between *agent* and *action*, or regards the absence of external impediments as common to the two states of freedom. (However, there is a difference here: in the case of deliberation the effectiveness of *any* external impediment is really always nil, since—to take the example arising in this context in Hobbes's later controversy with Bramhall—a man may *deliberate* whether or not to play tennis even if at the time of his deliberation the gates of the tennis court are locked, though of course, whatever his decision, he could never in actual fact have his game.) On the other hand, Hobbes's acceptance of the compatibility of freedom with determinism is perfectly defensible. Indeed it anticipates the position of 'compatibilism' or 'soft' determinism prominent among modern philosophers.

This concludes my account of the philosophically significant parts of Hobbes's *Critique* of White's *De Mundo*. There are many other topics on which I have not touched, for instance the issue of the vacuum, the origin of winds, comets, and the tides, the explanation of density and rarity, the nature of light, magnetism, gravity, the diurnal motions of the earth and moon, and question concerning eternity, fate, and astrology. My reasons for disregarding these aspects of Hobbes's polemic are (*a*) that most, if not all, have survived in detail in several of his mature writings, and (*b*) that on the whole they represent no more than antiquated thoughts in the history of science and philosophy. Throughout this early work, however, Hobbes's ideas and arguments reveal the acumen and pungency that one would expect from the author of *Leviathan* and the adversary of Bramhall.

I have noted two printing errors in the Latin text, *imaginetio* on p. 117, line 29, and *notus* for *motus* on p. 311, line 8. I also wonder whether the date given in Roman numerals on p. 47, line 15, should not be 1642 instead of 1652. I think that on p. 189, line 16, Mersenne's insertion of *cuba* is unwarranted, and that if a word is missing or wanted in this place it must be *cera*. I expect this point in the text will be rectified in H. W. Jones's forthcoming English translation.

W. VON LEYDEN
University of Durham

WASSERMANN, G. D. [1974]: *Brains and Reasoning*. London: Macmillan. £6·50. Pp. xi + 169.

The thesis of this book is that neuropsychology can shed a revealing light on central problems in the philosophy of science. In particular it seeks to refurbish Mill's theory about the psychological foundations of logic, replacing it by 'neuropsychologism' (p. 2). According to the author, inferences are only valid relative to the brains (or computers) which do the deducing (p. 14), and are themselves subordinate to thought processes. Philosophy has in general been led astray in failing to recognise its dependency on physiological considerations, the result of which has been an unwanted multiplication of problems and entities. Frege, for instance, made the mistake of believing 'that there exist formal relations which are "true" independently of the existence of human brains' (p. 17). Another mistake of traditional philosophy has been the distinction between what the author calls 'formal', and empirical statements, which ignores the fact that, 'at the neuropsychological level, recognition (classification) of both formal and empirical statements depends on previously formed memory traces . . .' (p. 16).

Now all these, admittedly programatic, remarks are open to many objections. First, it is certainly a mistake to think that, by substituting a well-developed neuropsychology for the vague psychology of Mill's time, one can provide an empirical foundation for logic. Psychologism in logic failed, not because the science of psychology was then too weak, but because it was realised that logic is a (normative) theory about the validity of inferences, not a descriptive account of the way in which people actually think, regardless of whether 'thinking' is interpreted psychologically or neurologically. To adapt an argument of Frege's, if we discovered beings whose brain processes were radically different from our own, to the extent that they regarded inferences of the form $p \vdash \neg p$ as valid, would we thereby have discovered a rival logic, or would logic remain the impartial arbiter of both theories of validity? Anyone who adopted the first alternative would have no right to ask the further question 'Who is correct?' In fact this question can be answered, since the distinction between valid and invalid inferences is based on considerations quite unrelated to brain processes. Again, it is not a very effective argument against Platonism to suggest that our conception of say, numbers, is brain dependent, since this clearly holds for our conception of empirical objects, the reality of which the author does not wish to deny. Finally, conventions, such as the distinction between empirical and analytic propositions may not be founded on the 'nature of things', but this will not impair their usefulness for certain purposes. The *value* of a distinction cannot be destroyed by an appeal to a common element in the origin of the two classifications.

Is there anything in the rest of this book to avoid these criticisms, and to illustrate the fruitfulness of the author's approach? It seems to me that there is not. Firstly, the author never substantiates any of his claims about neuropsychological logic. We never learn what it would be like, or how it would compare to its classical rival; but only that classical logic is in various ways 'inadequate'. For instance the author seems to think that the fact that theory construction is not 'a formal logical algorithmic procedure' is an example of the failure of a 'logical' approach to the sciences (p. 96). Nor do we learn how the traditional distinctions made by philosophers, *e.g.* that between analytic and synthetic propositions are *affected* by the insight that they are 'brain-dependent'.

Often the author is only restating in neuropsychological terms what philosophers have long known; *e.g.* that memory is not always veridical, that the objects that people see may conceivably be unreal (pp. 11–12). Also, the greater part of the book is taken up with the discussion of philosophical problems which the author does not even *attempt* to relate to neuropsychology. At these points his remarks generally consist either of rather ill-considered criticisms, or of objections that are already well known and to which he adds nothing. As an example of the former, Popper is criticised for believing that we can know when we have reached the truth (p. 32), whereas Popper has constantly stressed that we can never know any such thing. Examples of the latter abound: against operationalism, the author remarks that even empirical statements are theory-laden (p. 57); it is pointed out that falsifications are never conclusive (p. 92); realism is shown to be a metaphysical doctrine (p. 32), *etc.* In a text of less than one hundred and fifty pages almost every doctrine thrown up by twentieth-century philosophy of science gets some sort of mention. The result is that nothing gets adequately discussed.

So much for the purely philosophical pretensions of the book. But there is another side to the question. The ability to *recognise* validity, to make the distinction between empirical and non-empirical statements, and generally to carry on the activity of scientific enquiry, are surely interesting subjects for neurological investigation. (It is worth remarking that they presuppose, rather than imply, philosophical doctrines—we cannot investigate the mechanism of inductive learning until we distinguish inductive from deductive arguments.) Unfortunately the book fails even here. For instance the hopeful reader may turn to a section headed 'Thought Processes and Automata' (pp. 140–2), where he will learn only that humans are capable of creative thinking, while Turing machines are not; that human capacities vary widely; and that the best way to decide whether the brain is an automaton is to build automata that can match human abilities. Certainly, some issues which have traditionally been discussed by philosophers may benefit from an injection of empirical evidence from the neurosciences. 'Perception' and 'memory' are two good examples mentioned by the author, but they are not actually treated of in this book. Another philosophical problem that is usually thought of as being at least partly psychological is that of incommensurability. All the author says on this topic is that brains are capable of discovering similarities between non-identical concepts. This can hardly come as a surprise to anybody, though the author seems to think that it offers a complete solution to the problem (p. 140). As to the mechanism of inductive concept formation, the author's remarks are again disappointing: 'improved neuropsychological theories should be able to explain how brains can generalise' (p. 74).

Nor is the scientific information presented very well. As this book is aimed primarily at philosophers, one would expect there to be some explanation of the neuropsychological theories and evidence mentioned in it. There is none: they are baldly stated in the most uncompromising technical language, and we are referred to specialist journals for fuller explanations. This sort of esotericism can only serve to disguise the actual irrelevance of the material to the problems in hand.

This is a book in which I have found little merit. It is not well written, and the material, so much of which is anyway irrelevant, is ordered in an apparently arbitrary way. Above all, its arguments are vitiated by philosophical confusions and mistakes.

GREGORY CURRIE
London School of Economics

Brit. J. Phil. Sci. 26 (1975), 183–185 *Printed in Great Britain*

Recent Publications

BOOKS RECEIVED

Armstrong, W. (*ed.*) [1975]: *Francis Bacon: The Advancement of Learning, Book 1*. London: The Athlone Press. Hardback £4·00, Paperback £1·50. Pp. 153.

Appleton, I. (*ed.*) [1975]: *Leasure Research and Policy*. Edinburgh: Scottish Academic Press. £4·00. Pp. viii + 225.

Ayala, F. J. and Dobzhansky, T. (*eds.*) [1974]: *Studies in the Philosophy of Biology*. London: Macmillan Press Ltd. £12·00. Pp. vi + 390.

Bahm, A. J. [1974]: *Metaphysics*. New York: Barnes and Noble Books. Paperback: $2·95. Pp. vii + 259.

Bub, J. [1974]: *The Interpretation of Quantum Mechanics*. Dordrecht: D. Reidel Publishing Co. Pp. viii + 155.

Clavelin, M. [1974]: *The Natural Philosophy of Galileo—Essays on the Origins and Formation of Classical Mechanics*. London: The MIT Press. $25·00. Pp. xii + 498.

Cohen, R. S. and Schaffner, K. F. (*eds.*) [1974]: *P.S.A.* 1972, *Boston Studies in the Philosophy of Science*, **20**, Dordrecht: D. Reidel Publishing Co. Pp. viii + 445.

Cohen, R. S. and Seeger, R. J. (*eds.*) [1974]: *Boston Studies in the Philosophy of Science*, **21**, *Philosophical Foundations of Science*. Dordrecht: D. Reidel Publishing Co. Pp. vi + 545.

Cornides, T. [1974]: *Ordinale Deontik: Zusammenhänge zwischen Präferenztheorie, Normlogik und Rechtstheorie*. Vienna and New York: Springer-Verlag. $32·30. Pp. vi + 210.

Dangelmayer, S. [1974]: *Method und System: Wissenschaftsklassifikation bei Bacon, Hobbes und Locke*. Meisenheim am Glan: Verlag Anton Hain. Pp. 137.

Duncker, K. [1974]: *Zur Psychologie des Produktiven Denkens*. Berlin: Springer-Verlag. DM 32. Pp. iv + 135.

Elkana, Y. (*ed.*) [1974]: *The Interaction Between Science and Philosophy*. New Jersey: Humanities Press Inc. $17·50. Pp. ix + 481.

Enz, C. and Mehra, J. (*eds.*) [1974]: *Physical Reality and Mathematical Description*. Dordrecht: D. Reidel Publishing Co. Pp. viii + 552.

Fisk, M. [1974]: *Nature and Necessity*. London and Bloomington: Indiana University Press. £6·00. Pp. vii + 301.

Fitch, F. B. [1974]: *Elements of Combinatory Logic*. London and Cambridge, Massachusetts: Yale University Press. £3·50. Pp. vii + 162.

Foucault, M. [1974]: *The Archaeology of Knowledge*. London: Tavistock Publications. Paperback £1·80. Pp. 218.

Foucault, M. [1974]: *The Order of Things: An Archaeology of Human Sciences*. London: Tavistock Publications. Paperback £2·50. Pp. v + 387.

Gellner, E. [1974]: *The Devil in Modern Philosophy*. London: Routledge and Kegan Paul. £4·95. Pp. vi + 262.

Gellner, E. [1974]: *Legitimation of Belief*. London: Cambridge University Press. £4·25. Pp. 210.

Gonseth, F. [1974]: *Les Mathématiques et la Réalité.* Paris: Libraire Scientifique et Technique. Pp. ix + 386.

Gonseth, F. [1974]: *Les Fondements des Mathématiques.* Paris: Libraire Scientifique et Technique. Pp. vi + 243.

Gruber, H. E. [1974]: *Darwin on Man.* London: Wildwood House Ltd. £5·00. Pp. viii + 495.

Grunfeld, J. [1973]: *Science and Values.* Amsterdam: B. R. Grüner Publishing Co. Pp. 196.

Haack, S. [1974]: *Deviant Logic.* London: Cambridge University Press. £3·80. Pp. viii + 191.

Hamarneh, S. K. [1974]: *Ibn al-Quff.* Cairo: The Atlas Press. $6·50. Pp. vii + 199.

Harding, D. E. [1974]: *The Science of the 1st Person.* Ipswich: Sholland Publications. 75p. Pp. 48.

Hiorth, F. [1974]: *Noam Chomsky: Linguistics and Philosophy.* Oslo: Bergen-Tromsø-Universitetsforlaget. Pp. 10 + 180.

Hobbes, T. [1974]: *Traité de L'Homme.* Paris: Albert Blanchard. 30 F. Pp. 204.

Janich, P., Kambartel, F. and Mittelstrass, J. [1974]: *Wissenschaftstheorie als Wissenschaftskritik.* Frankfurt am Main: Aspekte Verlag GmbH. Dm. 14,80. Pp. 7 + 168.

Jardine, L. [1974]: *Francis Bacon: Discovery and the Art of Discourse.* London: Cambridge University Press. £4·90. Pp. viii + 267.

Holzhey, H. [1974]: *Wissenschaft/Wissenschaften.* Basel and Stuttgart: Schwabe & Co. Pp. 158.

Holzhey, H. [1974]: *Interdisziplinär.* Basel and Stuttgart: Schwabe & Co. Pp. 131.

Kopp, B. [1974]: *Beiträge zur Kulturphilosophie der deutschen Klassik.* Meisenheim am Glan: Verlag Anton Hain. Dm 29. Pp. 118.

Körner, S. (*ed.*) [1974]: *Practical Reason.* Oxford: Basil Blackwell. £5·00. Pp. viii + 264.

Leatherdale, W. H. [1975]: *The Role of Analogy, Model and Metaphor in Science.* Amsterdam: North-Holland Publishing Co. D.fl. 45.00. $17·50. Pp. 276.

Lehrer, K. [1974]: *Knowledge.* London: Clarendon Press. £4·65. Pp. viii + 256.

Levinas, E. [1973]: *The Theory of Intuition in Husserl's Phenomenology.* Evanston: Northwestern University Press. $7·50. Pp. xi + 163.

McDermott, F. E. (*ed.*) [1975]: *Self-determination in Social Work.* London: Routledge and Kegan Paul. Hardback £4·50, paperback £2·40. Pp. viii + 244.

Manuel, F. E. [1974]: *The Religion of Isaac Newton.* London: Clarendon Press/Oxford University Press. £3·50. Pp. 141.

Mini, P. [1974]: *Philosophy and Economics.* Florida: University of Florida Press. $8·50. Pp. 305.

Olby, R. [1974]: *The Path to the Double Helix.* London: Macmillan Press Ltd. £12·50. Pp. v + 510.

Peters, R. S. [1974]: *Psychology and Ethical Development.* London: George Allen & Unwin Ltd. £7·15. Pp. 480.

Quine, W. V. [1974]: *The Roots of Reference.* La Salle, Illinois: Open Court Publishing Co. $10·00. Pp. x + 151.

Radnitzky, G. [1974]: *Preconceptions in Research.* London: Literary Services and Production Ltd. 60p. Pp. 63.

Rapoport, A. [1974]: *Fights, Games and Debates.* Ann Arbor: University of Michigan Press. Hardback $7·95, paperback $4·95. Pp. viii + 400.

Rescher, N. [1974]: *Studies in Modality.* Oxford: Basil Blackwell. Pp. ix + 151.

Ritter, J. (*ed.*) [1971–74]: *Historisches Wörterbuch der Philosophie.* Basel and Stuttgart: Schwabe & Co. Dm 94.50. (First three volumes.)

Scheffler, I. [1974]: *Four Pragmatists.* London: Routledge & Kegan Paul. £4·75. Pp. vii + 269.

Schlesinger, G. [1974]: *Confirmation and Confirmability.* London: Clarendon Press/Oxford University Press. £2·65. Pp. 109.

Shapere, D. [1974]: *Galileo: A Philosophical Study.* Chicago and London: University of Chicago Press. Hardback £4·90, paperback £1·50. Pp. ix + 161.

Siu, R. G. H. [1974]: *Ch'i,a Neo-Taoist Approach to Life.* Cambridge, Massachusetts: The MIT Press. $10·00. Pp. 351.

Spilsbury, R. [1974]: *Providence Lost: A Critique of Darwinism.* London: Oxford Books. £3·50. Pp. xiii + 133.

Steffens, H. J. and Muller, H. N. (*eds.*) [1974]: *Science Technology and Culture.* New York: AMS Press Inc. $10·00. Pp. viii + 204.

Stevenson, L. [1974]: *Seven Theories of Human Nature.* London: Clarendon Press/Oxford University Press. Hardback £2·00, paperback 75p. Pp. 128.

Sutton, C. [1974]: *The German Tradition in Philosophy.* London: Weidenfeld & Nicolson. £4·00. Pp. v + 206.

Thomason, R. H. (*ed.*) [1974]: *Formal Philosophy: Selected Papers of Richard Montague.* London: Yale University Press. $12·50. Pp. 369.

Thorpe, W. H. [1975]: *Animal Nature and Human Nature.* London: Methuen & Co. Ltd. £7·20. Pp. v + 434.

Thrall, C. and Starr, J. (*eds.*) [1974]: *Technology, Power and Social Change.* Illinois: Southern Illinois University Press. $2·45. Pp. 169.

Will, F. L. [1974]: *Induction and Justification.* London and New York: Cornell University Press. £6·75. Pp. 371.

Wollheim, R. (*ed.*) [1974]: *Freud.* New York: Anchor Books. Pp. viii + 416.

Zimmerli, W. C. [1974]: *Wissenschaftskrise und Wissenschaftskritik.* Basel/Stuttgart: Schwabe & Co. Pp. 148.

BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE: ANNUAL CONFERENCE

The Society's 1975 Conference will be held at the University of Warwick on 26–28 September. The following sessions have been arranged:

Philosophy of Ethology: 'Explanation in ethology'

Main speaker: Dr A. Barker (Warwick)
Discussant: Dr M. Cullen (Oxford)
Chairman: Mr. V. Pratt (Cardiff)

Philosophy of C. S. Peirce: 'Regularity and explanation'

Main speaker: Mr R. Fellows (Bradford)
Discussant: Mr P. Gibbins (Hull)
Chairman:

Philosophy of Psychology: 'Problems of control and the functions of consciousness'

Main speaker: Dr D. Broadbent (Oxford)
Discussant: Dr T. Shallice (London)
Chairman: Professor J. Annett (Warwick)

Philosophy of Mathematics: 'The Vicious Circle Principle and Semantic Closure'

Main speaker: Mr G. G. Priest (St Andrews)
Discussant: Dr. D. Isaacson (Oxford)
Chairman:

The conference is being organised on behalf of the Society by Dr Susan Haack and Mr David Miller, both of the Department of Philosophy, University of Warwick, Coventry, Warwickshire, CV4 7AL, from whom further details may be obtained. (Please mark envelopes 'B.S.P.S. Conference'.)

Philosophy of Science

OFFICIAL JOURNAL OF THE PHILOSOPHY OF SCIENCE ASSOCIATION

Editorial Board

Richard S. Rudner (Editor-in-Chief), Peter Achinstein, Gustav Bergmann, May Brodbeck, Arthur Burks, C. West Churchman, Theodosius Dobzhansky, Adolf Grunbaum, Edward Madden, Henry Margenau, Gerald J. Massey, Alberto Pasquinelli, Wesley C. Salmon, Kenneth F. Schaffner, Dudley Shapere, J. J. C. Smart, Melford E. Spiro, Patrick C. Suppes, Lewis K. Zerby, Peter D. Asquith (Managing Editor). Book Reviews, James Van Evra. Assistant Editors, Patricia Ann Fleming and Paul C. L. Tang.

Vol. 40 December, 1974 No. 4

CONTENTS*

C. J. S. CLARKE **Quantum theory and cosmology**
NEAL GROSSMAN **The ignorance interpretation defended**
LARRY WRIGHT **Mechanisms and purposive behaviour III**
ERIC W. HOLMAN **Extensive Measurement without an Order Relation**
LOUIS NARENS **Measurement without Archimedean Axioms**
B. MAUND **Discussion: The Conventionality of Temporal Relations in Relativity Theory**
YORICK WILKS **Discussion: More on Fodor's Distinction between Strong and Weak Simulations**
A. ROSENBERG AND N. J. MACKINTOSH **Discussion: Strong, Weak and Functional Equivalence in Machine Simulation**

Book Reviews
Abstracts from *Inquiry*
Index to Vol. 41

*The contents of *Philosophy of Science* are indexed in *Science Citation Index, ABC Pol Sci, Language and Language Behaviour Abstracts*, and *The Philosopher's Index.*

PHILOSOPHY OF SCIENCE is published quarterly by the Philosophy of Science Association, 18 Morrill Hall, Department of Philosophy, Michigan State University, East Lansing, Michigan 48824. Subscription price $17.50 ($18.50 non-U.S.A.) per Volume. Members of the Association receive the Journal as a privilege of membership paid for out of membership dues. Inquiries about subscription or membership should be directed to the Managing Editor.

MIND

A Quarterly Review of Philosophy

edited by Professor **DAVID W. HAMLYN**

Vol. LXXXIV July 1975 No. 335

Articles	
Quine's Logic and the Class Paradox	P. SWIGGART
In Defence of the Predicate "Exists"	B. MILLER
Space and Analogy	M. GLOUBERMAN
Precision in History	J. J. KUPPERMAN
Justifying the Emotions	G. TAYLOR
The Right to Life	H. J. MCCLOSKEY
Discussions	
Has Harré Solved Hempel's Paradox?	N. GRIFFIN
D. F. Pears on "Is Existence a Predicate?"	D. A. GRIFFITHS
Resemblances and Universals: A Reply to J. Nammour	L. N. OAKLANDER
Causal Explanation and Human Action	E. CROWELL
On Preserving Entailment	E. W. JAMES
Classical Logic and Inexact Predicates—A Reply	S. KÖRNER
Critical Review	
Counterfactuals by D. Lewis	K. FINE

Books Received

Notice

Annual Subscription £2.50, single issues 80p

Published for MIND ASSOCIATION by

BASIL BLACKWELL 49 BROAD STREET OXFORD

INQUIRY An Interdisciplinary Journal of Philosophy and the Social Sciences
Edited by Arne Naess
and Alastair Hannay

Number 2 Summer 1975 Volume 18

Ecology in Ancient Greece	J. Donald Hughes
On the Human Question	John Rodman
'Science'. The Myth and its Role in Society	Paul Feyerabend
Why not Science for Anarchists too? A reply to Feyerabend	Arne Naess
Technological Society and its Counter-culture: An Hegelian Analysis	Clark Butler
From Apathy to Social Activism	Reuben Bitensky
Review Discussion:	
Ayer on the Problems of Philosophy (A. J. Ayer: *The Central Questions of Philosophy*)	Peter L. Heath

Published quarterly by: Universitetsforlaget, P.O. Box 307
Blindern, Oslo 3, Norway

Annual Subscription: Institutions – N. kr. 75.00, £4.90, $15.00
Individuals – N. kr. 60.00, £3.90, $12.00
Postage included

Studies in History and Philosophy of Science

VOLUME 6 NUMBER 1 APRIL 1975

CONTENTS

Editors Note

GARY THRANE: The Proper Object of Vision
NEAL WOOD: The Baconian Character of Locke's Essay

Notes on Contributors

Edited by Gerd Buchdahl and Mary Hesse Published quarterly by Pergamon Press

Annual Subscription Rates 1975
For libraries, research establishments and all other multiple-reader institutions: US$30.

Specially reduced rates to individuals. In the interest of maximizing the dissemination of material published in this important international journal we have established a two-tier price structure whereby individuals whose Institution takes out a library subscription, may purchase a second or additional subscription for their personal use at the reduced rate of US$15 per annum. All subscription enquiries should be addressed to: *Subscriptions Fulfilment Manager, Pergamon Press Limited, Headington Hill Hall, Oxford,* OX3 0BW.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by

NICHOLAS RESCHER

VOLUME 12/NUMBER 3 JULY 1975

CONTENTS

J. A. BROOK: *Personal Identity and Imagination*

HUGH S. CHANDLER: *Hedonism*

CARL WELLMAN: *Ethical Disagreement and Objective Truth*

TYLER BURGE: *Mass Terms, Count Nouns and Change*

JOEL J. KUPPERMAN: *Realism* vs. *Idealism*

LESTER HUNT: *Generosity*

ALEXANDER ROSENBERG: Propter Hoc, Ergo Post Hoc

PHILIP E. DEVINE: *The Perfect Island, the Devil and Existent Unicorns*

VOLUME 12/NUMBER 4 OCTOBER 1975

CONTENTS

ALAN SUSSMAN: *Mental Entities as Theoretical Entities*

MONROE C. BEARDSLEY: *Actions and Events: The Problem of Individuation*

DAVID PALMER: *Unfelt Pains*

J. A. BRUNTON: *The Devil is not a Fool*

CARL CRANOR: *Toward A Theory of Respect for Persons*

MARCUS HESTER: *Sensibility and Visual Acts*

WILLIAM DAVID SOLOMON: *Moral Reasons*

GEORGE GALE: *Are Some Aesthetic Judgments Empirically True?*

WILLIAM D. GEAN: *The Logical Connection Argument and* De Re *Necessity*

Annual Subscription: $20.00 to Institutions (*ex*cluding Annual Monograph, price $7.00)

$10.00 to Individuals (*in*cluding Annual Monograph)

Published by BASIL BLACKWELL

Orders to: 108 Cowley Road, Oxford OX4 1JF, England

The Journal of Philosophy

Subscriptions at $10.00/year; $7.00 to students
$12.00/year to libraries

Complete volumes and all separate issues available back to Jan. 7, 1904 (volume I, number 1). Prices as follows:

BEFORE 1963 (vols. I–LX)

Single issues: $1.50; double, $3.00. Volumes, $30.00 each.

1964 to 1973 (vols. LXI–LXX)

Single issues, $1.00; double, $2.00. Volumes, $24.00 each.

1974 AND LATER (vols. LXXI on)

Single issues, 60¢; double, $1.20. Volumes, $15.00 each.

ALSO AVAILABLE

Cumulative Fifty-year Index, 1904–1953; articles classified by subject and author; 452 p.; cloth, $12.00. Ten-year Supplement, 1954–1963: 98 p.; $3.00 cloth, $2.00 paperbound.

720 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NYC 10027

PHILOSOPHY DOCUMENTATION CENTER

Henri Bergson: A Bibliography. An up-to-date (through 1973) bibliography on the philosophic, literary, and scientific aspects of Bergson's thought. Annotated, 457 pages, hardbound. Published 1974.

Directory of American Philosophers. Handbook of philosophy in the United States and Canada. Faculties of philosophy, colleges and universities, fellowships and assistantships, societies, centers, institutes, journals and publishers. Biennial.

International Directory of Philosophy and Philosophers. World-wide handbook of philosophy (except U.S. and Canada). Faculties of philosophy, colleges and universities, societies, institutes, journals and publishers. Biennial.

The Philosopher's Index: An International Index to Philosophical Periodicals and the Philosophy Research Archives. Quarterly and annually. Subject and Author Index with Abstracts. Book Review Index. Translations in Progress. 260 journals indexed as well as papers of the Philosophy Research Archives.

Mailing Labels. 10,594 philosophers in the United States and Canada. Labels by philosophers' specialties, i.e., 399 in Aesthetics, 1,144 in Ethics, etc.

Philosophy Documentation Center
Bowling Green State University
Bowling Green, Ohio 43403 U.S.A.

Noûs

Nihil philosophici a nobis alienum putamus

A Quarterly Journal of Philosophy

edited by Hector-Neri Castañeda

Philosophy Department, Indiana University

Bloomington, Indiana 47401

NOÛS publishes quality essays and brief discussions on philosophical problems, regardless of the author's philosophical school or point of view. From time to time, NOÛS will publish symposia, surveys of work recently done on specially selected philosophical topics, critical studies of recent books (rather than reviews), and catalogues of recent publications in philosophy. NOÛS publishes the symposium papers of the A.P.A. Western Division meetings in its March issue.

A quarterly, NOÛS is published in March, May, September, and November each year. For *individuals* the subscription rates are: one year, $9.00; two years, $17.00; three years, $25.00. For *bona fide students:* $8.00 per year. For *institutions:* $16.00 per year (plus 10% for subscriptions handled through a subscription agency). Add 50 cents per year for all non-U.S. subscriptions. Make checks for one year subscriptions payable to INDIANA UNIVERSITY-NOÛS. Checks for multiple year subscriptions should be made payable to INDIANA UNIVERSITY FOUNDATION-NOÛS.

VOLUME IX, NUMBER 4, NOVEMBER 1975

CONTENTS

MICHAEL FRIEDMAN—*Physicalism and the Indeterminacy of Translation*

HARTRY FIELD—*Conventionalism and Instrumentalism in Semantics*

MARK PASTIN—*C. I. Lewis's Radical Foundationalism*

RICHARD B. ARNAUD—*A Note on 'Belief and Satisfaction'*

JOHN WALLACE—*Response to Arnaud*

Critical Study:

T. L. S. SPRIGGE—*Reinhardt Grossmann's* Ontological Reduction

Recent Publications in Philosophy

Order from

The Secretary, NOÛS

Department of Philosophy

Indiana University

Bloomington, Indiana 47401

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER

JUNE, 1975, VOLUME XXXV, No. 4

CONTENTS

I. ARTICLES
  Truth from Fiction?, by M. J. SIRRIDGE
  Fiction and Modality, by DENIS E. B. POLLARD
  Marx's Ethical Anthropology, by ALAN G. NASSER
  Materialism and Action, by N. M. L. NATHAN
  Four Varieties of the Logical Positivists' Doctrine of Necessary Truth, by CHARLES H. LAMBROS
II. DISCUSSIONS
III. COMMUNICATION, REVIEWS
IV. NOTES AND NEWS, RECENT PUBLICATIONS

Subscription prices $8.00 per year for individuals; $10.00 per year for institutions. Separate issues, $2.00 and $2.50 respectively. Postage and handling for back issues only. Papers submitted for publication will not be returned unless accompanied by a self-addressed stamped envelope, or equivalent postage.

A Circular listing the main contents of the Journal since 1940 will be sent upon request.

Address all communications to

PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH
State University of New York at Buffalo, Buffalo, New York, 14226

NOTRE DAME JOURNAL OF FORMAL LOGIC

BOLESLAW SOBOCINSKI, *Editor*
ANN W. LAUER, *Business Manager*

Published quarterly by the University of Notre Dame in January, April, July and October of each year, the Journal is devoted primarily to logic and the foundations of mathematics (i.e., algebraic logic, classical logical systems, constructive mathematics, methodology of deductive systems, model theory, non-classical formal systems, proof theory, recursion theory, and set theory).

Articles for publication and all editorial correspondence should be sent to the *Editor, P.O. Box* 28, *Notre Dame, Indiana* 46556, *U.S.A.*

Effective 1 July 1975 and beginning with Volume 17 (1976), subscriptions are $20.00 per year for institutions and $15.00 for individuals. Back volumes are $10.00 each for Volumes 1-9, $16.00 for Volumes 10-16. Orders for subscriptions and back volumes (accompanied by payment), and changes of address should be sent to the *Business Manager, P.O. Box 5, Notre Dame, Indiana* 46556, U.S.A.

7 NOV 1975

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1975-6

President: Professor D. J. O'Connor
Vice-President: Professor J. W. N. Watkins
Hon. Secretary: Helena Sheiham
Hon. Treasurer: Mr. D. Miller
Members: Mr. C. Howson, Dr M. Machover, Mr H. Martins, Mr. A. Young

EDITORIAL BOARD OF THE JOURNAL

Professor R. B. Braithwaite
Professor H. Dingle
Professor Sir John Eccles
Professor Sir Harold Jeffreys
Professor W. C. Kneale
Professor S. Körner
Professor M. Polanyi
Professor Sir Karl Popper
Professor G. J. Whitrow
Professor J. H. Woodger

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules :

' The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all persons who have been proposed by two members, approved by the Committee, and elected at an ordinary meeting. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £1·00 ($2.50). Those who wish to join should write to the Hon. Secretary, Helena Sheiham, Department of Philosophy, London School of Economics, Houghton Street, London WC2A 2AE.

Brit. J. Phil. Sci. Vol. 26 No. 2 June 1975

The British Journal for the Philosophy of Science

Contents

Articles

91–121 J. W. N. WATKINS Metaphysics and the Advancement of Science
123–132 RICHARD W. MILLER Propensity: Popper or Peirce?

Discussions

133–137 M. F. PODLAHA Some New Aspects of Relativity: Remarks on Keswani's Paper.
137–141 JAMES WOODWARD and WOLFGANG YOURGRAU Mach's Principle: Micro- or Macrophysical?

Review Articles

143–149 COLIN HOWSON The End of the Road for Inductive Logic? (review of R. Carnap and R. C. Jeffrey: *Studies in Inductive Logic and Probability*);
151–163 ALAN CODDINGTON Creaking Semaphore and Beyond (review of G. L. S. Shackle: *Epistemics and Economics*)

Reviews

165–182 JOHN BELL on A. A. Fraenkel, Y. Bar Hillel and A. Levy: *Foundations of Set Theory;* M. F. HALLETT on Hao Wang: *From Mathematics to Philosophy*; W. VON LEYDEN on Thomas Hobbes: *Critique du* De Mundo *de Thomas White*; GREGORY CURRIE on G. D. Wasserman: *Brains and Reasoning.*

Recent Publications

Programme of Meetings

7 NOV 1975

Printed in Great Britain at the University Press, Aberdeen

Volume 26 Number 3 September 197

BJPIA5 26 (No. 3) 187-274 (1975) ISSN 0007-0882

The British Journal for the Philosophy of Science

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editors: JOHN WATKINS and JOHN WORRALL

Assistant Editors: COLIN HOWSON and ELIE ZAHAR

Subscriptions

The *Journal* is published quarterly. The subscription price is £7.00 net (U.S.A. and Canada U.S. $22.00) per volume post free; single issues £2.50 net (U.S.A. and Canada U.S. $8.00) plus postage. The subscription is £4.50 (U.S.A. and Canada U.S. $14.00) for members of the British Society for the Philosophy of Science, who should order direct from the publisher. (Membership of the Society is international and is by no means confined to people resident in the United Kingdom.) Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

Notes to Contributors

Articles, correspondence and books for review should be sent to John Worrall at the Department of Philosophy, London School of Economics, Houghton Street, Aldwych, London WC2A 2AE. Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Footnotes, which should be as few as possible, should be typed separately, double-spaced at the end of the text. They should be referred to by consecutive numbering throughout the text, in which they will be inserted in page proof. Bibliographical references should be listed alphabetically at the end. The style of references should follow current *Journal* practice.

Long quotations are printed in small type without quotes, and should be so marked in the typescript. Single quotes should be used to form names of words and for shorter quotations; double quotes for other purposes. Diagrams, tables, and illustrations should be on separate sheets, with their desired position in the text clearly indicated. Symbolic formulae appearing as separate lines should be clearly distinguished in the text, and foreign letters (Greek, German, etc.) indicated in the margin. Material to be printed in italics, especially English letters used as symbols ('*A*', '*x*', etc.), should be underlined. On matters of style and layout, authors should consult CHAUNDY, T. W. *et al.* [1964]: *The Printing of Mathematics*, London: Oxford University Press.

Offprints of articles may be ordered when corrected galley proofs are returned. Fifty are supplied free ; the price of extra copies is given on the order form.

© British Society for the Philosophy of Science 1975

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editors.

Back Volumes

Volumes 1-16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17-25 may be obtained from Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Prices on application.

Brit. J. Phil. Sci. 26 (1975), 187–198 *Printed in Great Britain*

The Rule of Succession, Inductive Logic, and Probability Logic

by COLIN HOWSON

Since the later seventeenth century there has existed a theory of inductive probability, which attained an impressive mathematical development at the hands of Laplace; in the nineteenth century it was identified by its advocates as a branch of logic, dealing with a type of inference ostensibly a generalisation of the deductive variety. A fundamental principle of the new logic was the principle of indifference,[1] or insufficient reason.[2] This principle may be stated thus: if there is no reason relative to our information to suppose that among a class of events none is any more likely than any other, then their probabilities are equal, relative to that information. This principle, combined with the additivity and normalisation rules for probability functions, generates the so-called classical definition of probability over the set of events definable within a finite possibility space of n elements, where all the n elementary events are indifferently favoured, so to speak, by the conditions defining that space. In Laplace's hands, the principle of indifference seemed to generate certain probability assignments sensitive to purely numerical features of the outcomes of a repeated experiment. An example is the formula called by John Venn the Rule of Succession:

$$P(e_{i+1}/e_1 \wedge \ldots \wedge e_i) = \frac{r+1}{i+2}, \qquad r \leqslant i$$

where e_j, $j = 1, \ldots, i+1$ record the outcomes of $i+1$ experiments to determine the presence or absence at each of a given empirical character, where 'e_{i+1}' asserts that the character will be manifested at the $i+1$th trial, and '$e_1 \wedge \ldots \wedge e_i$' asserts that it is present in r out of i trials. The probability function is thus a strictly increasing function of r for constant i, which tends to 1 in the limit for $r = i$ as i tends to infinity.

This application of the principle of indifference does not—unlike some others[3]—seem to be such as is likely to invite formal inconsistencies (I shall substantiate this claim later in this section of this paper). Yet it is precisely because it generates—*ex nihilo*, as it were—such justifications of induction as are expressed in the celebrated posterior probability assignments like the Rule of Succession, that it has been looked upon with increasing suspicion

Received 15 *October* 1974

---

[1] Keynes [1921], p. 44.

[2] Von Kries [1886], p. 6.

[3] See the list in Keynes [1921], pp. 44–54.

as to its formal soundness, from the early nineteenth century onwards. Indeed, there is an impressive body of twentieth century opinion[1] unanimous in declaring the principle formally unsound because the Rule of Succession is itself contradictory. If successful, this charge would eliminate at a stroke one of the best known and certainly mathematically most developed justifications of enumerative induction; if unsuccessful, our attempts to argue against this type of induction, would have somehow, it seems, to show directly that the principle of indifference is, even in local, well-determined contexts like sequences of repeated binomial experiments, unsound. I propose, however, in this paper, to show that the charge of inconsistency brought against the Rule of Succession is unsuccessful—indeed, that the charge is based upon a simple fallacy—and to defend the principle of indifference (in a more precisely formulated form) as a valid, though trivial, principle of logic. In so doing I shall argue that properly interpreted it does not yield the Rule of Succession nor indeed any inductive principle. In the third and last section I shall sketch a plausible and promising development of the logical interpretation of probability.

1 The charge of inconsistency brought by the distinguished writers cited above is easily rebutted by examining the assumptions from which the Rule of Succession follows and showing that they do not in fact deliver the type of contradiction (which I shall shortly mention) which it is alleged that they do.

The Rule of Succession is a result about a type of random sampling. Let us consider m disjoint and exhaustive empirical properties $A_1, \ldots A_m$ of individuals presumed to be the possible qualitatively distinct outcomes of a type of repeatable experiment, which in principle can be repeated indefinitely. The space of all possible indefinitely repeated experiments of this type is just $\{A_1, \ldots A_m\}^N$, where N is the set of natural numbers. Now the Rule can be derived either as an approximate, *i.e.* limiting formula where we replace N by a variable finite number n (see, for example, Jeffreys [1939], p. 127), or explicitly for the infinite case, which, incidentally, Laplace himself considered. It is more verisimilar, it seems to me, to consider this case, *i.e.* to consider the set of possible trials as infinite.[2] Now we shall want to discuss the convergence of sequences of random variables defined on the space Ω, where these random variables are certain 'sample variables'; accordingly the class of events relative to which our probabilities are defined will include the σ-field generated by the finitely indexed cylinder subsets of Ω. In fact, we might as well take this field itself as our field of events. Let us assume that we have no reason to believe that the appearance of any of the characters at one experiment influences its or any other character's appearance at subsequent trials; this being the case the probability measure is naturally symmetric[1] (this is the first application of the principle of indifference). Let us, to simplify matters, also suppose that we are interested only in observing the character A_i and that the probability measure is symmetric with respect only to the two derivative characters A_i, $\bar{A}_i$. Let us accordingly define a sequence $x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)} \ldots$ on Ω such that $x_k^{(i)}$ takes the value 1 on a member of Ω if the kth element is A_i, and 0 if not. Define $y_n^{(i)} = \frac{1}{n}\sum_{j=1}^{n} x_j^{(i)}$. By a theorem of de Finetti

---

[1] Keynes [1921], p. 412, Broad [1968], p. 90, Kneale [1949], p. 204, Carnap [1950], p. 568.

[2] It is traditional to represent this set as the set of balls contained in an infinite urn, the characters $A \ldots, A_m$ being, for example, distinct colours. It was with such a concrete model in mind that Laplace approached the mathematical problem; and Karl Pearson was wont to refer to the universe as a 'Nature Urn': the set Ω of infinite sequences then represents all possible constitutions of the urn. The mathematical reasoning does not, however, require such a model, and I mention it only as a historical curiosity.

$$P(x_1^{(i)} = 1, x_2^{(i)} = 1, \ldots, x_n^{(i)} = 1) = \int_0^1 y^n \, dF(y)$$

where $F(y)$ is a distribution function and $F(0) = 0$ and $F(y) = 1$ for $y > 1$. $F(y)$ is the limiting distribution function of a sequence of a step functions F_n which define the distributions of the variables $y_n^{(i)}$. Following Laplace we set F_n to be the step function with a step of height $1/(n+1)$ at each of the points $0, 1/n, 2/n, \ldots, 1$.[2] Then $F(y) = y$ and

$$P\left(y_n^{(i)} = \frac{r}{n}\right) = \binom{n}{r} \int_0^1 y^r (1-y)^{n-r} \, dy = \frac{1}{n+1}$$

Hence

$$P\left(x_{n+1}^{(i)} = 1 \mid y_n^{(i)} = \frac{r}{n}\right) = \frac{\int_0^1 y^{r+1} (1-y)^{n-r} \, dy}{\int_0^1 y^r (1-y)^{n-r} \, dy} = \frac{r+1}{n+2} \quad \text{—(1)}$$

This is the Rule of Succession. The second application of the principle of indifference is seen in the definition of the functions F_n; they express an assumption that we have no reason to believe that any of the possible frequencies $0, 1, \ldots, n$ of A_i-type characters is any more likely than any other in finite samples.

The charge of inconsistency is based on the following assumption. If the Rule of Succession holds relative to the character A_i, then it holds relative to any other. Suppose that $m > 2$. By the additivity property of P the probability that the $n+1$th trial will exemplify the character A_1 or A_2

---

[1] In other words $P(B') = P(B)$ where B' is obtained from B by a finite permutation.

or . . . or A_m, conditional on a sample in which K_1 results are of the type A_1; K_2 of the type A_2, *etc.*, is 1. But then the Rule of Succession gives that same disjunctive event the probability

$$\sum_{j=1}^{m} \frac{(K_j+1)}{n+2} = \frac{n+m}{n+2} \neq 1$$

But this is nonsense; for the Rule, *i.e.* the formula (1) is derived relative to a particular family of binomial variables $x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}, \ldots$. If we want to consider the sample as a multinomial sample, on the other hand, represented by the values of a multinomial variable family x_k defined analogously to the $x_2^{(i)}$ except that $x_k = 1$ if A_1 is the kth element, $x_k = 2$ if A_2 is the kth element, *etc.*, then the analogue of (1) is

$$\frac{\int_0^1 \int_0^{1-x_1} \ldots \int_0^{1-x_1-x_2-\ldots-x_{m-2}} x_1^{K_1+1} x_2^{K_2} \ldots x_{m-1}^{K_{m-1}} (1-x_1-\ldots-x_{m-1})^{K_m}\, \mathrm{d}x_1 \ldots \mathrm{d}x_{m-1}}{\int_0^1 \int_0^{1-x_1} \ldots \int_0^{1-x_1-x_2\ldots-x_{m-2}} x_1^{K_1} x_2^{K_2} \ldots x_{m-1}^{K_{m-1}} (1-x_1-\ldots-x_{m-1})^{K_m}\, \mathrm{d}x_1 \ldots \mathrm{d}x_{m-1}}$$

giving the probability that the next trial will give the character A_1, where $K_1+K_2+\ldots+K_m = n$, as $(K_1+1)/(m+n)$. Clearly the probability that the $(n+1)$th trial yields the disjunctive character A_1 or $A_2 \ldots$ or A_m is 1, as desired.

What then of the customary diagnosis of the 'contradiction', squarely blaming an unrestricted use of the principle of indifference?[1] First of all it has not produced a contradiction here because there is no contradiction. But its acquittal at this tribunal should not be interpreted as an invitation for it to resume the position it once enjoyed as a quite general method of discovering probabilities; in particular it most certainly can lead to absurdities. To take an example of R. A. Fisher's, we may have an equally distributed ignorance about the values that will be taken by two random variables; but if one is obtained by a non-linear transformation of the other they cannot both be uniformly distributed (assuming even that with probability one their range of values is covered by a finite interval). Nevertheless we feel very strongly that a symmetrical stochastic experiment should generate a set of outcome points—where that set is measurable—over which probability is uniformly distributed; and very often we are right. Occasionally it is possible to provide a mathematical argument to tell us why we were right: for example, the theory of Markov chains provides a fairly simple justification for the fact that the more a pack is shuffled, the more approximately equal are the probabilities of the distinct hands that might be dealt[2]; the same theory explains why shaking an urn containing black and white balls many times results in approximately equal pro-

---

[1] Broad [1968], p. 90; Kneale [1949], p. 204.

[2] *Cf.* Feller [1957], p. 407.

babilities for the possible states of the urn (this example incidentally furnishes a model for the diffusion of two gases); Poincaré showed that the probability that a freely revolving pointer will stop in a segment of arc length δx tends—as the number of rotations within a circle of unit radius increases—to the value characteristic of a uniform distribution, $\delta x/2\pi$: and so on.

But there are notable exceptions to uniformity, relative to the variable we think should be uniformly distributed. An obvious example arises in micro-physics where Maxwell–Boltzmann statistics for distributions in phase space give way either to Bose–Einstein or Fermi–Dirac statistics, contrary to entrenched intuition. It follows that while uniform distributions nevertheless remain well represented in physics (and of course genetics), the uniformly distributed variable is often not the one determined by our intuition of how the symmertical set up should behave. To sum up, we cannot evade the difficulty represented by the possibility of non-linear transformations of random variables by attempting to adduce empirical support for the contention that just one of these is 'naturally' distributed uniformly; this being so the principle of indifference as a general heuristic principle is certainly unsound.

But I think it possible—though this is an unfashionable task—to redeem the principle, in its intended interpretation, which was as a principle of logic. When it is so interpreted it is—I claim—a theorem of (meta) logic, though some of its ostensible consequences—like the Rule of Succession—are not. An examination of this interesting, some have thought pathological, area of human reason suggests that the Rule of Succession in particular, and 'inductive principles' in general, are not logical, confirming the well known claim of Sir Karl Popper.

2 Let us consider an urn containing n black and white balls in an unspecified proportion. A language to describe states of the urn might contain as its only extralogical terms n distinct individual constants $a_1, a_2, \ldots, a_n$ and one one-place predicate symbol P. Our set up is described by the sentence

$$A \ldots \bigwedge_{1<i<j\leqslant n} a_i \neq a_j \wedge \forall x[\bigvee_{i\leqslant n} x = a_i]$$

Now the number of distinct canonical interpretations of the language satisfying the conjunction of A and a 'state description' attributing P to some subset of the set of constants and $\overline{P}$ to the remainder is just one, where the classes of models (whose domains are, say, subsets of the natural numbers[1]) are factored by isomorphism. There are 2^n such conjunctions containing A and it is plausible to define the degree to which one of the state descriptions is entailed by A as the quantity $1/2^n$,[1] *i.e.* the proportion of models of A which are also models of the state-description. All the usual formulas of the theory of chances for drawing from the urn without replacement are clearly derivable by conditioning on the disjunction of state descriptions with the appropriate constant number of instances of P, conjoined with A. In the case corresponding to 'complete ignorance' of the proportion of black balls, the measure is obtained, of course, by summing over a product of what I shall from now on call probabilities, according to the calculus of probabilities. Where the proportion of black to white balls has not been given, has been the traditional province of the principle of indifference, and has accordingly been viewed with growing suspicion from the early nineteenth century. (The first case has usually been regarded as within the domain of a frequency account, with the condition being extracted from the probability functor and placed as a condition defining the class of repeatable trials). But both cases arise perfectly naturally within this theory of generalised (logical) derivation, the difference between them lying only in how much information is premissed, *i.e.* included in the condition; both cases, therefore, exemplify the principle of indifference, stated in the objective mode thus: if the dependence of one statement on a consistent set of premisses is the same as that of all other statements within a given finite class, their (conditional) probabilities are equal.[2] But this is trivially true (it cannot be extended to denumerably infinite sets, for the obvious reason). In general the ratios of measures of the classes of models cannot be defined without arbitrariness where the set of conditioning statements is satisfied in infinite domains,[3] but we shall concentrate here on the cases where these sets have only finite (and hence bounded) models, for it is with respect to these that the problems in the theory of chances which I want to examine are chiefly defined.

---

[1] They must be restricted to sets, and this seems a perfectly natural way of restricting their size.

It is possible by these considerations to explain the widespread doubts about the application of the principle of indifference in the Bayes–Laplace theory of inductive probabilities, including the nexus of formulas connected with the Rule of Succession among which are those defining the famous Laplacean posterior distributions. There appears to be no way, within the

[1] *Cf.* Wittgenstein [1961] proposition 5.142, p. 81; though the credit for first having defined such a logical measure is Bolzano's; unlike his theory of deductive logic, his theory of partial inference has not borne much fruit, largely because of an apparently inescapable arbitrariness of which he was well aware, relative to conditions not limited to finite domains (Bolzano [1837], section 161].

[2] von Kries put the principle thus 'two hypotheses are equally probable if they comprehend equal, indifferent and original *Spielräume*' (*indifferente ursprüngliche Spielräume*): von Kries [1886], p. 157. But for him the *Spielraum* was simply the range of possibilities of an indefinite proposition.

[3] *Cf.* Howson [1973].

theory I have briefly sketched, of combining the property of symmetry with equal weights given to the various distinct proportions of white balls that are possible, from 0 to n. If our language has names for all the individuals then the principle of indifference appears in the equal weights assigned to the canonical models, of which there are 2^n (satisfying the sentence A, of course). If on the other hand its only extralogical term is the predicate symbol P then the sentence A is transformed into

$$A' \ldots \qquad \exists x_1 \ldots \exists x_n \left[\bigwedge_{i<j\leqslant n} (x_i \neq x_j) \wedge \forall x \left[\bigvee_{i\leqslant n} x = x_i \right]\right]$$

and the conjunction of A' with the sentence σ_m asserting that there are exactly m instances of the predicate P has one model up to isomorphism. Hence A' has $n+1$ models, the sentences σ_m all having probability $1/n+1$. But the 'sample sentence' asserting that R out of N individuals have P, where $N \leqslant n$, has as many distinct models as there are distinct integer solutions of

$$x+y = n-N \qquad\qquad x, y \geqslant 0$$

namely $n-N+1$, Hence such a sentence has a probability given A' which is independent of R, namely $n-N+1/n+1$.[1] It follows from these reflections that the Rule of Succession is not of a logical nature, whatever else it might be. It also follows that only one value in Carnap's celebrated $\lambda = \infty$, or $\mathbf{m}^{\dagger}$—Wittgenstein's measure. Maxwell–Boltzmann statistics, of course, correspond to the case where our language includes names; to the case where it does not, correspond the Bose–Einstein values. Nature, in favouring the latter, seems to have decreed naming to be more than a human artifice.

It is easy to show that if a finiteness condition is included, probabilistic logic is not inductive, even in the limit, contrary to some strong advocacy from the eighteenth century onwards into the first half of this century. We really have two cases, corresponding to our two types of language above (the intermediate case is no different as far as results go, but it presents an unnatural compromise which I shall not consider). Let us regard all probabilities as conditioned on A, respectively A', and a variable set of other conditions. In the first case, relative to A and any m-fold sample, the probability that a_{m+1} will have the predicate P is exactly $\frac{1}{2}$; the measure here makes the random variables x_i, $i = 1, \ldots, n$, x_i taking the value 1 on $P(a_i)$ and 0 on $\overline{P}(a_i)$, independent.

In the second case (*i.e.* our language without names) relative to A′ and any n-fold sample reporting that there are just R out of $N < n$ instances of P observed, the probability of at least one further instantiation of P is independent of R, and equal to $(n-N)/(n-N+1)$, which tends to 1 as $n \to \infty$, for fixed N. The probability that all the remaining $n-N$ observations will instantiate P is $1/(n-N+1)$, which is also independent of R and tends to 0 as $n \to \infty$. Inductive logic, it seems is not logic.

---

[1] That the probability values depend on which criterion we adopt to distinguish between models does not make them in the least arbitrary; the obvious criterion is that we should regard as indiscernible models which cannot be distinguished within the language relative to which the probabilities are defined. The sets of models in each of the above cases are isomorphic to the Lindenbaum sentence algebra of the languages.

It is my opinion that the traditional theory of chances was an attempt to articulate such logical relations as these, and that the plausibility of the characteristically Laplacean methods, which propelled the tissue of confusion known as inductive logic on its way, was due to the lack of a developed logical notation by which to distinguish the correct from the incorrect 'applications' of the principle of indifference.[1] It is tempting to draw an analogy between the principle of indifference and the theory of truth: both generate contradictions when the languages to which they are applied are in a certain way ill-defined. To this extent the impressive list of contradictions Keynes draws up under the principle of indifference is similar to, say, the paradox of the Liar. With the aid of a developed notation both types of contradiction disappear. In the infinite case, of course, the principle of indifference does *not* generate a unique distribution, as Bertrand's celebrated paradoxes of 'geometrical' probability demonstrate: here the choice of a random variable (in deciding by which types of co-ordinates a chord of a circle might be represented, to use Bertrand's own example) determines the distribution.

3 This theory is applicable only to the negligible class of inferences whose premisses are satisfied only in bounded (finite) domains, and even then involves only the mathematics of proper fractions, and it is understandable in view of its weakness that it should be neglected. Furthermore, although we can prove inequalities like Chebychev's relative to a sequence of languages successively augmented by a new constant, their 'predictions' about the successive sample means are not to be taken seriously in the practical context of a binomial experiment; for those experiments are practically certain to be characterised by a binomial parameter which differs from $\frac{1}{2}$.

Nevertheless the idea of a probability function measuring—in some manner—the dependence of each event in a field of events upon the conditions defining a stochastic experiment, or in the actuarial domain, a reference class, is attractive, and makes one wonder whether such measures of dependence can be defined relative to conditions introducing asymmetries, like loaded dice, biased coins *etc.*—the rocks on which the theory of chances ran aground. In general, of course, a distribution function is a function—often a continuous one—not only of a set of random variables but also of a set of parameters $\theta_1, \ldots, \theta_k$ characterising the relevant features of the experiment: the interpretation of such functions that I should like to propose is that where $P(X) = p = p(\theta_1, \ldots, \theta_k)$, the *dependence* of X on the conditions of the experiment is actually measured by max $(p, 1-p)$. X is determined (more accurately, 'almost determined') by those conditions, then, if it has a probability of 0 or 1 relative to them. But the outcome of a stochastic experiment has, of course, determinants other than the experimental conditions; for clearly, if the fundamental structure of the universe were sufficiently different, no doubt we should encounter different probability distributions (as quantum theory implies that the classical distributions of the kinetic theory are not everywhere even approximately accurate). So we can tentatively suggest that the form and particular parameter values of a distribution function reflect the joint influence of an (unknown) set $\mathscr{L}$ of universal laws and a set of initial conditions provided by the experiment.[1] It is plausible that any natural scale for the measurement of this 'quantity' can be recalibrated to the closed unit interval, the measure being transformed into a probability measure.[2] Now any probability measure can be regarded as a quotient of measures: for we can set $P(A) = \mu(A \cap B)/\mu(B)$ where B is the full space of outcomes and μ is proportional to P. If an experiment is defined by a description $\mathfrak{S}$, including a definition of the outcome space, then B can be regarded as the set (suitably reduced) of models of $\mathfrak{S} \cup \mathscr{L}$, and A as the measurable subset of these satisfying the description of a given event. Hence we obtain a probability function P which can be interpreted as a ratio of measures on classes of models. If the probability of A relative to the given type of experiment is p, then we can represent the dependence measure max $(p, 1-p)$ as the degree to which A is *causally dependent on the experimental conditions.*

---

[1] Bolzano and von Kries were not the only philosophers in the nineteenth century to conceive of a logical interpretation of the probability calculus in this way that I have sketched. Sigwart, in his [1895] writes 'the theory of this branch of inference from disjunctive propositions has not, however, been developed in a form corresponding as this does to the ordinary logical mode of representation; it has taken shape as the mathematical theory of probabilities, which is shorter and more elegant, at any rate in more complicated cases, and which is fundamentally nothing but a mathematically formulated chapter from logic' (p. 219).

---

[1] This seems more natural than interposing shadowy 'propensities', as Popper does, which contribute nothing by way of explanatory power and if anything reduce it (see points (i)–(v) below). Popper himself once advanced the thesis that these propensities are to be 'intuitively understood' as measures of causal influence; but he subsequently ignored, it in favour of 'real' propensities (Popper [1966], p. 39).

[2] I. J. Good was among the first probabilists to point out how general were the conditions under which a real valued set function could be transformed into a probability measure (Good [1948], p. 105). He uses a theorem of Abel's, to the effect that if a real valued function $f(x, y)$ of two real arguments x and y is commutative and associative then $f(x, y) = \phi^{-1}(\phi(x)+\phi(y))$ for some ϕ. A rather similar theorem was proved by Aczél (Aczél [1948]).

Let us suppose that we have an n-fold repeated experiment, and on the product space determined by it is defined a sequence $x_1, \ldots, x_n$ of binomial random variables. The conditional distribution of any subset of these, relative to the remainder taking values in a certain interval, measures, in this theory, the causal influence (modulo $\mathscr{L}$) of the conditions defining the original experiment jointly with that of the conditions expressed in the second argument of the conditional probability function. If the addition of these new conditions makes no difference to the distribution then the outcomes of this augmented experiment are causally independent of those extra conditions. Thus, assuming that the events determined by the value of the random variables avoid the extreme probability values, we can identify probabilistic independence with causal independence in the logical sense in which that term is used here.

By a familiar argument from the compactness theorem for first order logic we can show that our probability functions are continuous on their field of definition. This permits us to do two things: (a) to discuss arbitrary sets of sentences, for there is a unique extension of a probability function on the Borel closure of that field; and (b) to make use of the various theorems of mathematical probability which concern countably additive probability functions. This second fact enables us to show that relative to a given experiment there is an 'almost unique' measure; for by Borel's theorem the probability measure relative to any fixed set of conditions delivering on repetition independent outcomes is such that $P(X)$, where X is an event defined relative to one trial, is equal to the limiting relative frequency of X except on a set of measure 0. There is consequently little arbitrariness about the choice of probability measure, though it is, of course, not *a priori* computable.

This account explains the following:

(*i*) The prediction that *any* set of conditions whose repeated instantiations give causally independent trials usually generates a class of relative frequencies for each event which are stable, *i.e.* they remain within bounds $(p-\epsilon, p+\epsilon)$ for fixed p after sufficiently many trials, and approximately constant from sequence of repetitions to sequence of repetitions. (Notice that it is plausible to say that a *particular* statistic is explained, when its probability is shown to be greater than $1-\delta$, though it was not, of course, predictable: thus the symmetry between explanation and prediction is characteristic of just the fully determinate case).

(*ii*) The dependence of a distribution upon a set of parameters representing certain features of the experiment.

(*iii*) The possibility of systematic bias in experiments, characterised by the systematic satisfaction by a sequence of experiments of conditions additional to those defining each experiment, and hence leading to the 'wrong' distribution relative to those conditions.

(*iv*) The possibility of explaining a distribution by reference to a set of deterministic laws ($\mathscr{L}$). The explanation is a weak explanation, of course, not in general amounting to complete, but rather to incomplete, entailment. This theory also explains the compatibility of random events (*i.e.* those obeying a product distribution law) with a locally deterministic environment—for example, that containing such macrophysical phenomena as dicing, coin-tossing, *etc.*

No other theory of probability, to my knowledge, has as great an explanatory power. The modern frequency theory inaugurated by Kolmogorov[1] has an almost purely operationalistic definition of probability as long run relative frequency (though with the saving clause that this is manifested only with 'practical certainty'); Popper's propensity theory does appeal to real quasi-forces, propensities, but these cannot be empirically characterised independently of frequencies, and certainly (*i*) above is not a straightforward prediction of this theory.

It seems to me that interpretative theories of probability must satisfy the following methodological canon: they should explain, and the more the better. To shift the meaning of the probability function so that it reflects a certain metaphysical taste, and no more, is not good enough, particularly as there are enough real problems to be solved in the philosophical theory of probability. I think that the above account satisfies our canon well enough.

The theory here presented is a fairly straightforward development of the logical theory of probability due to Bolzano. It has the merit of unifying the sundered interpretations of the probability calculus; for the subjective accounts are explained as viewing logic, and not merely probability, subjectively; physical probabilities are explained as the empirical manifestation of the component $\mathscr{L}$ in the measure of the dependence of the description of a trial outcome on an (incomplete[2]) set of 'initial conditions' and the set of (non-statistical) laws $\mathscr{L}$[3].

Finally, we can note that a problem of Waismann's can now be solved.[4] He is considering a die, and notes that if it is not loaded it shows six, say, in one sixth approximately of all of a large enough sequence of throws, and that the frequencies respond in an appropriate way to shifts of the centre of gravity. The frequency theory, he writes, does not and cannot 'link up the laws of deterministic physics with statistical fact. . . . We all feel that there is some kind of link' and 'it may be convenient to postpone the examination of an inexplicable problem for a time, but in the long run it can never be simply dismissed'. Waismann proposes a solution similar to ours, which we can round off nicely by observing that if the die is symmetrical about its centre of gravity, its outcomes are independent of classical dynamics, or indeed any physical laws. For if we take a language containing arbitrary non-logical terms, in which the sample space is describable, and we condition *only* on this description, then each outcome is assigned an equal probability, which must, by the calculus of probabilities, be 1/6. If we now incorporate in the condition the laws of dynamics as an exact enough approximation then, on the grounds that the probability measure p is 'almost everywhere' equal to the limiting relative frequency, and that the relative frequency in finite samples is a consistent estimator of p, we set the value of p at approximately 1/6. The values are (approximately) the same in both cases and so the outcomes are independent of the extra information.

---

[1] Kolmogorov [1933], p. 4. Also *cf.* Cramer [1946], p. 148.

[2] Thus the success of probability theory is absolutely no support for indeterministic metaphysics, any more than is the existence of incomplete theories in general.

[3] These laws may be very incomplete, even as far as conditional predictions are concerned.

[4] Waismann [1930], pp. 230, 231.

London School of Economics

REFERENCES

Aczél, J. [1948]: 'Sur les opérations définies pour nombres réels', *Bulletin de la Société Mathematique Française*, **76**.

Bolzano, B. [1837]: *Wissenschaftslehre.*

Broad, C. D. [1968]: *Induction, Probability and Causation.*

Carnap, R. [1950]: *Logical Foundations of Probability.*

Cramer, H. [1946]: *Mathematical Methods of Statistics.*

Feller, W. [1957]: *An Introduction to Probability Theory and Its Applications.*

Good, I. J. [1948]: *Probability and the Weighing of Evidence.*

Howson, C. [1973]: 'Must the logical Probability of Laws be Zero?', *British Journal for the Philosophy Of Science*, **24**, pp. 153–63.

Jeffreys, H. [1939]: *The Theory of Probability.*

Keynes, J. M. [1921]: *A Treatise on Probability.*

Kneale, W. C. [1948]: *Probability and Induction.*

Kolmogorov, A. N. [1933]: *Foundations of the Theory of Probability.*

Kries, J. von. [1886]: *Die Principien der Wahrscheinlichkeitsrechnung.*

Popper, K. R. [1966]: 'Quantum Mechanics without "The Observer"' in M. Bunge (*ed.*): *Quantum Theory and Reality.*

Sigwart, C. [1895]: *Logic* (English translation of *Logik*).

Waismann, F. [1930]: 'Logische Analyse des Wahrscheinlichkeitsbegriffs', *Erkenntnis*, **1**.

Wittgenstein, L. [1961]: *Tractatus Logico-Philosophicus* (Pears-MacGuiness translation).

Brit. J. Phil. Sci. 26 (1975), 199–211 *Printed in Great Britain*

Grünbaum on the Metric of Space and Time

by PAUL HORWICH

1 This paper is about Adolf Grünbaum's theory of the magnitudes of spatial and temporal intervals. It is based upon his many writings on the subject which have recently been collected together in the Second Edition of his *Philosophical Problems of Space and Time.*[1]

Grünbaum's key idea is that facts about the equality (congruence) of intervals are not solely spatial or temporal facts, but concern the relations between space or time and certain physical standards such as rulers and clocks. In his words, the metrics of space and time are extrinsic. This view leads Grünbaum to conventionalism. It follows, he believes, that whether or not two intervals are congruent is a matter of convention.

I shall try to show first of all that Grünbaum provides only a vacuous basis for his so-called conventionalism. Secondly, I discuss his view that the metrics of space and time are extrinsic. This claim is criticised as regards (*a*) its justification and plausibility (*b*) its compatibility with Grünbaum's causal theory of time, and (*c*) its claim to be an elucidation of Riemann's position. Finally I take up Grünbaum's criticism of Geometrodynamics—the theory that matter is made out of space-time. I argue that the difficulty rests upon an equivocation, and I attempt to clarify and resolve the issue by means of Kripke's notion of a reference fixing definition.

2 Grünbaum's brand of conventionalism about space and time must be carefully distinguished from the conventionalist theses of those who have written in the spirit of Mach, Poincaré and Reichenbach. This has been emphasized by Fine,[2] Glymour,[3] and Friedman[4] and by Grünbaum himself. For whereas, for example, Reichenbach, writing in the tradition of positivism, bases his conventionalist claims upon epistemological features of space and time, Grünbaum explicitly disavows these roots and instead founds his conventionalism upon certain alleged ontological properties of the space, time and space-time manifolds. Reichenbach's

Received 10 *August* 1974

[1] Grünbaum [1973]. My exposition is largely based on his [1970], reprinted in his [1973].
[2] Fine [1971]. [3] Glymour [1972]. [4] Friedman [1972].

view in a nutshell is simply this:[5] that since there is in principle no way to check whether some process, taken to be uniformly periodic, is *really* uniformly periodic, our assumption that it is (for which there is no empirical backing) must therefore rest upon the adoption of a convention to that effect. Thus Reichenbach's conventionalism depends upon the impossibility of verifying that our congruence standards (and the theories which embody them) are indeed correct. Or, in other words, it depends upon the possibility of finding alternate congruence standards, and correspondingly alternate theories, which would be perfectly adequate to render the observed facts. The differences between these standards and associated theories would reside in their relative complexity. Thus we may be well motivated to adopt some particular standard because of the relative simplicity of the physical theories which are engendered by this choice. However, for Reichenbach such pragmatic reasons for the conventional adoption of a particular congruence standard do not impugn the fact that the choice is conventional. The other standards and theories would be just as correct and true, and this is sufficient to make our particular selection a matter of convention.

Grünbaum's claims about spatial and temporal congruence are not based upon such verificationist considerations. Rather, his views derive merely from the presumed fact that space, time and space-time are *continuous* manifolds and from a consequence of this fact; namely, that these manifolds are *intrinsically metrically amorphous.* Before going on to explain this notion, I want to illustrate the radical difference between Grünbaum and Reichenbach. It turns out on Grünbaum's view that discrete space is not intrinsically metrically amorphous. And he holds, as a consequence of this, that, despite the possibility of alternate metrics for discrete spaces, those metrics would not hold simply as a matter of convention. Thus, unlike Reichenbach, Grünbaum thinks that alternate metrisability is not a sufficient condition for the conventionality of any particular metric. This is because there could be a number of different ways to metrise a manifold in terms of the same intrinsic properties of its elements. In such a case there could be alternate metrics alright, but each one, when it applied, would apply as a matter of intrinsic-to-the-manifold fact.

But what does Grünbaum mean by 'an intrinsic property' and by the notion of 'intrinsic metric amorphousness'? And, more importantly, why should these ideas be relevant to the issue of conventionalism as it has traditionally been understood as opposed to the special way in which Grünbaum construes it?

[5] Reichenbach [1927].

Grünbaum says that the inspiration for his view is to be found in the following sentences from Riemann's Inaugural Dissertation:

> Determinate parts of a manifold, distinguished by a mark or by a boundary, are called quanta. Their comparison as to quantity comes in discrete magnitudes by counting, in continuous magnitude by measurement. Measuring consists in superposition of the magnitudes to be compared; for measurement there is requisite some means of carrying forward one magnitude as a measure for the other. In default of this, one can compare two magnitudes only when the one is part of the other, and even then one can only decide upon the question of more and less, not upon the question of how many (much).
>
> . . . in the question concerning the ultimate basis of relations of size in space . . . the above remark is applicable, namely that while in a discrete manifold the principle of metric relations is implicit in the notion of this manifold, it must come from somewhere else in the case of a continuous manifold. Either then the actual things forming the groundwork of a space must constitute a discrete manifold, or else the basis of metric relations must be sought for outside that actuality, in colligating forces that operate upon it.[1]

Thus Grünbaum's conventionalism may be cast into Riemann's terminology, as follows. When the metrical properties of a manifold are implicit in the manifold (as in the case of discrete manifolds) then those properties hold—whenever they do hold—as a matter of *solely* spatial (or temporal) fact. When, however the metrical properties ascribed to elements of the manifold are based upon circumstances outside—not implicit within it—then those properties are not strictly genuine spatial properties. Although they hold as a matter of some facts or other, those facts are not solely spatial facts. Therefore *qua* spatial properties they are convention-laden since the alternative congruence definitions are not merely providing us with equivalent ways of rendering one and the same set of pre-existing spatial facts, but rather consist in alternative ways of imposing metrical structure upon a manifold which does not possess one in its own right.

At this point two questions naturally arise. First, how can we tell which features of a manifold are *implicit* within it and which are not? And in particular how do we know that the metrical properties of space, time and space-time are not implicit in those manifolds? To answer this we must take account of Grünbaum's further explication of the notion of 'implicitness'. This is where 'intrinsic properties' and 'intrinsic metrical amorphousness' come in. Second, why should we be concerned about Grünbaum's conclusion, given that it is stated in terms of a notion of 'conventional' which is idiosyncratic? As we have seen, in Grünbaum's sense, the notion does not connect up in the usual way with the possibility of choosing between equally good alternative standard or theories. So we may wonder what, if anything, of philosophical significance is being said when Grünbaum asserts that the metrics of continuous space, time and space-time are convention-laden.

---

[1] Riemann [1854].

3 I shall consider the second question first. It seems to me that the idiosyncratic use of 'conventional' in Grünbaum's claim that the metric of space is conventional *because* space is intrinsically metrically amorphous, renders that claim vacuous. In the first place, Grünbaum nowhere presents an argument for this claim. He says merely that this basis for conventionalism gives the theory a significance which it would not otherwise possess. For if conventionalism is based simply upon the possibility of alternative congruence standards—that is, upon the possibility of giving alternative interpretations, in the model-theoretic sense, to certain theoretical terms—then, on his view, conventionalism becomes trivialised since we can always find alternative interpretations for the terms of any uninterpreted formal calculus. Thus conventionalism would no longer be a theory which had any special relevance to space and time. Even if this is granted, it still does not show that Grünbaum has given any content to the claim that the metric is conventional *because* it is not intrinsic. All it shows is the obvious fact that we convey more information when we say that the metric is conventional *and* not intrinsic than we do when we say merely that the metric is conventional. Grünbaum's claim about the connection between intrinsicality and conventionality is, it seems to me, made vacuously true by his decision to restrict the term conventional to only those properties which are not intrinsic. And I think that my objection is supported by the observation that Grünbaum apparently does not feel the necessity to argue for the thesis that the metric of space is conventional *because* metrical properties are not intrinsic to space.

There is a further reason to think that this claim lacks content. Consider why it is that we are concerned about whether the statements in some domain are true as a matter of convention. Often the motivation derives from our concern over the epistemological status of the statements. How do we know which of them are true; what justifies our knowledge? When no factual (for positivists, observational) backing presents itself to justify some knowledge claim, then it has been thought plausible to suppose that we know that the statement in question is true because we have adopted certain conventions which guarantee it. The statement is held to be true, not as a matter of fact, but rather as a matter of convention. But now consider again Grünbaum's claim. Is it really *because* metrical properties are not intrinsic to space that they apply to certain spatial intervals as a matter of convention? The answer is no, since, as we have seen, the extent to which a congruence claim is backed by facts is independant of whether those facts are intrinsically spatial or not. Even though it may be admitted that the facts are not solely spatial, there are nonetheless some facts or other—facts, say, about the relation between two intervals and a transported meter-stick—which (once the standard has been prescribed) provide the empirical backing for such congruence claims. In other words, the extent to which the truth of a congruence statement is conventional in an epistemologically interesting way depends upon the extent of its factual backing in general, and not upon its backing in intrinsically spatial facts. Insofar as Grünbaum restricts his application of the term 'conventional' to those properties which are non-intrinsic, he departs from that use of the term which made it philosophically interesting. Of course one may always stipulate a more restricted use. But then the extra significance of its application is just a trivial reflection of these further requirements. In the same way I might give extra significance to the word 'book' by stipulating that nothing with less than 200 pages could count as a book. Then of course when I say that something is a book I have conveyed more information than usual. This is not to deny that it may well be philosophically interesting and important that spatial and temporal metrics are not intrinsic. But it is to deny then there is any content to the claim that those features provided a basis for their conventionality.

4 I now want to turn back to our first question and examine Grünbaum's notion of the 'inherence' of a property within a manifold. This is of importance, not only for its relevance (if any) to conventionalism, but also for its bearing upon Grünbaum's criticism of the Clifford–Wheeler programme of geometrodynamics. Moreover the question of the ontological status of spatial and temporal metrical properties is interesting in its own right.

Grünbaum's account is a generalised explication of Riemann's distinction between metric relations which are 'implicit in the notion of a manifold' and metric relations 'whose principle must come from somewhere else' and 'whose basis must be sought for outside that actuality (manifold) in colligating forces that operate upon it'. It consists essentially of the following definitions concerning any given manifold M.

(1) A entity is *internal* to an interval of M *iff* existence of the interval depends on the existence of the entity. Thus every element belonging to an internal (a, b) is internal to (a, b) in this sense.
(2) A monadic property P is *external* to an interval possessing it *iff* the obtaining of P depends on entities which are not internal to the interval.
(3) A monadic property P is *intrinsic* to an interval possessing it *iff* P is a general property (one whose constitution does not involve particular individuals) of the interval and not external to it.

(4) A partition $\prod$ of the class of intervals of a manifold M is *intrinsic* to the intervals of M *iff* $\prod$ is generated by an equivalence relation that is intrinsic to the intervals of M.

(5) A metric L is intrinsic to the intervals of M *iff* the partition generated by the congruence relation is a partition intrinsic to the intervals of M.

Analogous definitions are framed to cover polyadic properties and weakened versions are formulated to capture the notion of things being internal (external, intrinsic) to the manifold M, rather than to the intervals of the manifold. So the general idea is that a congruence relation is intrinsic to a spatial or temporal manifold just in case the actual congruence of any two intervals of the manifold does not imply (in the sense of logical entailment) the existence of any entity whose existence is not already implied by the existence of the manifold.

Thus, according to Grünbaum, the property of being one meter long is not intrinsic to continuous space, since an interval has that property only in virtue of the existence of some non-spatial standard such as a meter stick. However that property would be intrinsic to discrete space since in that case a meter would correspond to a certain number of space-atoms; and the cardinality of the set of space-atoms in an interval is in intrinsic property of the manifold, since its obtaining does not imply the existence of anything other than space itself. Similarly, the relation of temporal betweenness is held to be an intrinsic property of the temporal continuum since it is held that the particular order in which instants occur does not require the existence of anything that is not already required by the mere existence of the temporal manifold.

There is a further problem, which Grünbaum sees as an epistemic difficulty, and which arises from the fact that several different predicates may denote one and the same property. We may have some way of characterising a congruence relation, which indicates that the relation is not intrinsic to a certain manifold. But how can we know that there is not some other way to describe exactly the same relation—some way which would indicate that the relation is in fact intrinsic to the manifold? According to Grünbaum if we do have several descriptions of a relation, all of which indicate that it is intrinsic to some manifold, then we should take this as inductive evidence that the relation is in fact extrinsic. What a description D indicates about the intrinsicality or extrinsicality of the congruence relation or metric it describes is, Grünbaum says, 'a matter of individual mentality with respect to what D suggests to a particular person'.

5 There seem to me to be several problems with Grünbaum's account.

(*A*) As an explication of Riemann's views it seems dubious. For when Riemann says: 'while in a discrete manifold the principle of metric relations is implicit in the notion of the manifold, it must come from somewhere else in the case of a continuous manifold', it seems plain that he means to indicate an *essential* difference between discrete and continuous manifolds. For why else should he say that in a discrete manifold the principle of metric relations is implicit in the *notion* of the manifold. And that in the case of a continuous manifold it *must* come from somewhere else. Reimann does not restrict himself to the manifolds of space and time, and he obviously intends to maintain that, of necessity, any continuous manifold lacks an implicit principle of metric relations. Yet Grünbaum surprisingly construes Riemann's claim about continuous space as an empirical hypothesis. Not only does Grünbaum maintain this interpretation explicitly, but also his own commitment to this view of continuous space is evident in our above exegesis. For he states clearly that, even on the assumption that space is continuous, we can have at best only strong inductive evidence that there exists no inherent basis for metric relations. He calls this Riemann's Metrical Hypothesis. Thus Grünbaum appears to believe that we just *might* discover that continuous space does in fact possess an intrinsic metric. And this, rather than constituting an explication of Riemann's view, seems straightforwardly incompatible with it.

(*B*) On Grünbaum's account, our evidence that space (time) has no intrinsic metric consists in our evidence that space is continuous together with the fact that all descriptions of possible metrics have so far indicated that they are extrinsic. This latter fact is held to count as good inductive evidence that the metrics characterised by those descriptions really are extrinsic. Now this seems unsatisfactory in two respects. In the first place, whether a description D of a metric indicates that the metric is extrinsic is held to be 'a matter of individual mentality with respect to what D suggest to a particular person'. Thus there are, on this account, no objective criteria according to which we may establish (or get evidence for) the intrinsic or extrinsic status of a metric. Must we conclude that Grünbaum's views on the status of spatial and temporal metrics are not susceptible to rational debate? Secondly, the assumption that we may acquire inductive evidence about the intrinsicality of the metric of space by examination of all the different predicates we can find which denote metrical properties, is not at all obviously true. It gives the surely mistaken impression that our inductive evidence is linguistic. Further clarification of the inference is required. Moreover it behoves Grünbaum to specify at least roughly what kind of experimental or theoretical results would lead him to conclude that the metric of our continuous space is actually intrinsic. Without this information it is difficult to follow let along evaluate the inductive inference which allegedly produces Riemann's Metrical Hypothesis.

(*C*) Insofar as one subscribes to an absolutist conception of ontologically autonomous space (time or space-time) it is prima facie possible to follow Grünbaum and distinguish those properties of the manifold which are intrinsic to it from those properties which its elements may possess only in virtue of the existence of entities distinct from the manifold. If, however, one denies, with the relativist, the autonomous existence of the manifold, then such a distinction may become problematic.

Such a difficulty arises for Grünbaum in his treatment of time. On the one hand he maintains that the metric of time is extrinsic to time. That is to say he views the obtaining of congruence between pairs of temporal intervals as dependent upon the existence of entities (such as clocks) distinct from time itself. On the other hand he holds to a version of the causal theory of time. This denies the autonomous existence of time. Temporal betweenness is reducted to a set of causal relations between events, leaving an ontology of events in space between which certain relations obtain.

Now the following specific problem arises for Grünbaum. He cannot maintain the extrinsicality of the temporal metric merely on the grounds that congruence involves the existence of physical entities such as events. This is insufficient reason, since, according to the causal theory, the mere existence of time also implies the existence of events. Thus the congruence of intervals implies nothing whose existence is not already implied by the existence of time.

If the extrinsicality of temporal congruence is to be demonstrated then it must be shown that the sort of entities required for the obtaining of congruence *i.e.* clocks—are distinct from the sort of entities required by the causal theory for the obtaining of temporal betweenness.

It might be thought that this is a simple matter. Is it not obvious that there could be worlds with enough events in them, causally related in the right way, to generate the existence of time, yet no clocks—no periodic processes which could serve as an intrinsic basis for a temporal metric? The answer is no. It is true that only in certain 'nice' cases will there be processes whose designation as periodic would lead to simple laws of nature. However it is Grünbaum's own view that any string of events may be specified as defining a periodic process. Thus Grünbaum's view of the metric of time appears to be incompatible with his causal theory of time.

(*D*) In ascribing a metric to space we commit ourselves, according to Grünbaum, to the existence of entities whose existence is not required just by the existence of space. This is why the metric of space is said not to be intrinsic to space. It follows that it is at least logically possible that there be a space with no metric. For if there could be no such thing, then the postulation of a space with a metric would, contrary to Grünbaum's view, involve no more ontological commitment than the mere postulation of space.

Consider such a space with no metric and consider two disjoint intervals within it, *AB* and *CD*. What can be said about the relationship between these intervals? Since there is no metric it should certainly not be the case that *AB* and *CD* are congruent. But let us suppose that if, contrary to fact, there had been a rigid rod in this world exactly coincident with *AB*, then it could have been transported so as to be exactly coincident with *CD*. Such a supposition raises a problem for Grünbaum. For the counterfactual relation ascribed to *AB* and *CD* clearly provides the basis for the metrical claim that *AB* and *BC* are congruent. However, that basis is intrinsic to space since it does not imply the actual existence of any non-spatial entity.

Grünbaum might reply to this objection as follows. The counterfactual property is not intrinsic to space, since the predicate makes reference to rigid rods. However, this reply is to no avail.

In the first place; whether or not the predicate used to denote the property makes mention of any nonspatial entities is beside the point. According to Grünbaum's definitions a general property is intrinsic to space just in case its obtaining does not depend upon entities upon which the existence of space does not depend. In other words, a general property is intrinsic to space just in case there is no entity *x* such that the existence of space does not logically entail the existence of *x*, whereas the obtaining of the property does entail the existence of *x*. The counterfactual, dyadic property, 'if there were a rigid rod coincident with ——, it could be transported so as to be coincident with ——', satisfies this condition. Its obtaining does not imply the existence of any actual rigid rods; nor does it imply the existence of anything other than space. Therefore, since it is a general property, it is intrinsic to space.

Secondly, even the hint of a reference to rigid rods may be eliminated. Whenever an object has a counterfactual property, it has some categorical property in virtue of which the counterfactual claim is true. For example, if an object would have dissolved if it had been placed in water, then this is because of some categorical property of the object—its chemical constitution. In our case, the categorical property is that of being equal in length. It is because *AB* and *CD* are equal in length that the rigid rod would behave in the way described. Neither the obtaining of this counterfactual property, nor that of its categorical basis—the congruence relation—implies the existence of non-spatial entities. Moreover, the predicates 'is congruent to' and 'is equal in length to' do not even make mention of non-spatial entities.

6 I think that the above points, especially the last one, constitute grave difficulties for Grünbaum's account. It seems to me that this last problem derives from an assumption that the *meaning* of metrical relations is to be explained in terms of the relations between intervals and physical standards If this view is adopted then Riemann's intuitions are satisfied and the essential features of Grünbaum's theory appear plausible. However, there is a good alternative to this view of the matter—an alternative indicated by Kripke[1] which would account perfectly well for Riemann's observations, yet which would avoid the difficulties of Grünbaum's position. The alternative is to hold, not that mention of external standards is essential in the explanation of the *meaning* of metrical relations. But rather, they are required to *fix the reference* of metrical relations. On this view the assertion of a metrical property or relation no more commits us to the existence of external standards than does an attribution of yellowness commit us to the existence of the recipients of yellow sensations. We may fix the reference of the term 'yellow' by the reference-fixing definition, 'Yellowness causes persons to have sensations of yellow' without committing ourselves in every attribution of yellowness, to the existence of persons. And in a similar way we may fix the reference of spatial and temporal congruence relations without committing ourselves in every attribution of those relations to the existence of the standards whose use was essential in the reference fixing definitions. Thus, Riemann's claims are preserved on this account. It *is* necessary that we have recourse to some things outside the spatial and temporal manifolds. They are required to pick out the metrical relations which obtain nonetheless without them.

7 This approach suggests a way of replying to Grünbaum's critique of geometrodynamics. Grünbaum argues that there is a fundamental difficulty with the Clifford–Wheeler programme of geometrodynamics.

> For Clifford tells us . . . that matter and radiation are indeed *constituted* by curved space. And it would certainly *seem* that *in that case*, the very curvature of space *cannot* itself depend on any 'non-implicit' metrization by transported bodies. Instead, the curvature would need to obtain with respect to a metric implicit in empty space . . . But such a denial of *RMH* (Riemann's Metrical Hypothesis) is *inconsistent* with Clifford's claim that it is 'wholly devoid of meaning' to ask whether the lengths of our familiar measuring standards 'do really change by mere moving about, without our knowing it'.

The problem, in other words, seems to be this. On the one hand, the very essence of geometrodynamics is the thesis that objects are constructions out of the metrical properties of space: *i.e.* that the metrical properties of space are ontologically primitive with respect to matter and radiation.

[1] Kripke [1971].

Yet, on the other hand, it is claimed by the proponents of geometrodynamics (correctly, according to Grünbaum) that the metric of space is not ontologically autonomous but rather must be defined in terms of relations of coincidence between objects: *i.e.* that objects and such relations between them are ontologically primitive with respect to the metrical properties of space. Thus conventionalism and geometrodynamics appear to be incompatible with one another.

I think that there are two aspects to the solution of this problem.

In the first place, the senses in which matter is, and is not, *ontologically primitive* with respect to the metrical properties of space, are different. Thus the apparent contradiction rests upon an equivocation. For, according to geometrodynamics, matter is build out of the metrical properties of space in just the way that atoms are built out of electrons, protons, *etc.* and in the way that macroscopic objects are built out of atoms. That is the sort of notion of *construction* upon which is founded the claim that the metric of space is ontologically primitive with respect to matter. From conventionalism, however, we learn that space has no intrinsic metric—that space derives its metrical properties from our conventional ascription of self-congruence to certain physical objects selected as 'rigid under transport'. On this view, objects are held to be primitive in that it is to statements about them that statements about the metric of space must be reduced. Thus both matter and the metrical properties of space are each ontologically more primitive that the other—but in different respects. Metrical properties are primitive in the sense that they constitute the blocks from which matter and radiation are built, and matter is primitive in the sense that the attribution to space of metrical properties depends essentially upon some specification of the behaviour of matter or radiation.

These mutual relations of primitiveness are not incompatible with one another. For in a similar vein an operationalist may claim (1) that physical objects are made out of subatomic particles, and yet (2) that statements about such entities derive their meaning from (are reducible to, should be analysed as) statements ascribing observable properties to macroscopic objects. It may be that Grünbaum's conventionalism is a far cry from operationalism. However that does not detract from the anology. For the point of it is merely to show that the two senses of being ontologically primitive are such that two sorts of entity may, in different respects, each be primitive with respect to the other.

The second part of my approach to Grünbaum's difficulty relies upon Kripke's distinction between definitions which give the meaning of an expression and definitions which merely fix the reference in the actual world of an expression. Thus on Kripke's view the definition 'A meter is the length of rod S' merely fixes the reference of the term 'meter'. It does not presume to specify the meaning of the term, since there are possible worlds in which it is false that S is a meter long. In some worlds the rod is bent, in some it does not even exist, and yet the term would nonetheless apply to other rods in those worlds. Now if it is supposed, as we have claimed above, that the connection between the metrical properties of space and the behaviour of matter is achieved by means of reference fixing definitions, then Grünbaum's problem dissolves. For there is nothing paradoxical in the facts that (1) one sort of entity, *A*, may be constituted out of another sort, *B*, and (2) reference to the entities, *B*, must be fixed by definitions framed in terms of entities, *A*.

According to Riemann and Grünbaum the metrical relation between two spatial intervals *AB* and *CD* is not implicit in the space manifold. Recourse must be taken to some external congruence standard. Thus we may, as a matter of convention stipulate that rods of a certain sort maintain their length under transport, and we may, as a matter of fact discover that some such rod coincides with the interval *AB* and subsequently with *CD*. It is in virtue of this mixture of convention and fact that *AB* and *CD* would be held to be congruent. However we need *not* say that such hypothetical procedures give an account of the meaning of the congruence claim. Rather we may say that they constitute a reference fixing definition. They provide us with a way of (partially) fixing the reference of the metrical term 'congruent'. Now having conventionally fixed the references of metrical terms (by means of the description of such procedures involving physical objects) we may proceed with the programme of geometrodynamics and perform the theoretical reduction of matter to metrical properties in terms of the metric which has already been conventionally circumscribed. As we have seen, this approach conforms to Reimann's view that reference to external standards are necessary for the ascription of metrical properties to space. At the same time since it eschews the idea that the attribution of metric is tantamount in *meaning* to the ascription of properties to these standards, it therefore permits the programme of geometrodynamics to escape from the charge that any space theory of matter could amount to no more than a matter theory of matter.

Massachusetts Institute of Technology

REFERENCES

Fine, A. [1971]: 'Reflections on a Relational Theory of Space', *Synthese*, **22**, pp. 448–481.

Friedman, M. [1972]: 'Grünbaum on the Conventionality of Geometry', *Synthese*, **24**, pp. 219–35.

GLYMOUR, C. [1972]: 'Physics by Convention', *Philosophy of Science*, **39**, pp. 322–40.

GRÜNBAUM, A. [1970]: 'Space, Time and Falsifiability', *Philosophy of Science*, **37**, pp. 469–588.

GRÜNBAUM, A. [1973]: *Philosophical Problems of Space and Time. Boston Studies in the Philosophy of Science*, **12**. Dordrecht: D. Reidel.

KRIPKE, S. [1971]: 'Naming and Necessity', in D. Davidson and G. Harman (*eds.*): *The Semantics of Natural Language*. Dordrecht: D. Reidel.

REICHENBACH, H. [1927]: *The Philosophy of Space and Time*. New York: Dover, 1958.

RIEMANN, B. [1854]: 'On the Hypotheses Which Lie at the Foundation of Geometry', in D. E. Smith (*ed.*): *A Source Book in Mathematics*, **2**. New York: Dover, 1959.

Brit. J. Phil. Sci. 26 (1975), 213–225 *Printed in Great Britain*

Putnam and Clarke and Mind and Body

by YORICK WILKS

Introduction

1 *Putnam's Thesis.*
2 *Putnam's Analogy.*
3 *Abstract and Real Machines.*
4 *Putnam's Counter Attack.*
5 *Clarke's Arguments.*
6 *Clarke's Arguments about Machines are Misconceived.*
7 *Clarke and Putnam.*
8 *Privacy and 'Turing's Test'.*

INTRODUCTION

The analogy between human beings and machines has been philosophically fruitful since Descartes, and continues to bear in our own time. Machines themselves have changed utterly during that period, and it is therefore all the more striking that what is perhaps the most interesting exploitation of the analogy in recent years should be in terms of a definition of 'machine' so primitive that it could have been stated and understood at any time during the last two millennia. I am referring to Putnam's ingenious analogy[1] between a Turing machine and the human body/mind. Putnam sought to show that the body/mind distinction could be suitably defined for a Turing machine (*TM* for short), and argued that, since we would not want to take the distinction too seriously in the case of a, trivially simple, *TM*, there was no need to do so in the case of a person either, because the very same arguments were used to establish the distinction in the two entities.

In this paper I want to do three things: firstly, to attack the basis of Putnam's analogy, by arguing that it rests on a systematic equivocation over the notion of the 'state' of a *TM*. Secondly, to consider some recent arguments of Clarke in this *Journal*,[2] also directed against Putnam but which, in my view, do not tell against his position. Thirdly, to make, but not develop in detail, a frivolous speculation about the possible privacy of a machine.

Both Putnam and Clarke, in different ways, entertain this last notion. Putnam, as an *advocatus diaboli*, constructs a trivial privacy for a trivial

Received 8 *July* 1974

---

[1] Putnam [1961] developed further in his [1964] and [1967].

[2] Clarke [1972].

machine, and argues that, to accept human privacy is also, necessarily, to accept this trivial form. Clarke, implicitly, develops the notion of privacy for a machine more like those we meet in modern laboratories of computation or banks, but he in no way wishes to defend it. I, unlike both Putnam and Clarke, postulate a machine privacy, and for machines of an only slightly fanciful sort, which I claim should be taken a little more seriously. Moreover, if I am right that Putnam's analogy with human privacy does not hold, then from the machine privacy I discuss nothing at all disquieting follows about human beings.

I PUTNAM'S THESIS

Putnam argues that

> . . . all of the issues [that make up the traditional mind-body problem] arise in connection with any computing system capable of answering questions about its own structure and have thus nothing to do with the unique nature (if it is unique) of human subjective experience.[1]

Putnam then describes such a system, a *TM*, and inevitably, in any discussion of Putnam's argument, the *TM* must be redescribed. The precise form of the description is not important, but I will give an abbreviated version of Putnam's own sketch:

> . . . a *TM* is a device with a finite number of internal configurations, each of which involves the machine's being in one of a finite number of STATES, and the machine's scanning a tape in which certain symbols appear. The machine's tape is divided into separate squares . . . on each of which a symbol, from a fixed alphabet, may be printed. Also the machine has a 'scanner' which 'scans' one square of the tape at a time. Finally, the machine has a printing mechanism which may (*A*) erase the symbol which appears on the square being scanned, and (*B*) print some other symbol (from the machine's alphabet) on that square.
>
> Any *TM* is completely described by a MACHINE TABLE, which is constructed as follows: the rows of the table correspond to [symbols] of the alphabet (including the 'null' letter, *i.e.* blank space), while the columns correspond to states *A*, *B*, *C*, *etc.* In each square there appears an 'instruction', *e.g.* '*s*5*LA*' . . . [which] means 'print the symbol *s*5 on the square you are now scanning . . . and proceed to scan the square immediately to the left of the one you have just been scanning; also, shift into state *A*'.
>
> A 'machine table' describes a machine if the machine has internal states corresponding to the columns of the table, and if it 'obeys' the instructions in the table in the following sense: when it is scanning a square on which the symbol *s*1 appears, and it is in, say, state *B*, that it carries out the 'instruction' in the appropriate row and column of the table (in this case column *B* and row *s*1).[2]

The *TM* is set in its initial state, with an initial symbol in the square to be scanned first, and its machine table will then 'run' it until it stops in a rest state. The symbols printed out on the tape in the course of the running are normally interpreted as the result of the computations that the machine table defines.

---

[1] Putnam [1964], p. 138.

[2] Putnam [1964], pp. 141–2.

It is important to note that the *TM*, as described, is an abstract entity that could be realised physically in many different ways. The *TM* as described is not a machine at all, in any real sense, since it necessarily requires someone or something to read the machine table. This distinction between an abstract and a concrete *TM* is fundamental to any criticism of Putnam's thesis and analogy.

2 PUTNAM'S ANALOGY

Putnam's analogy hinges on the relation between two statements, one made 'by' a *TM* and one by a person:

(*i*) I am in state *A* if, and only if, flip flop 36 is on,

where flip flop 36 is some switch in a real *TM* that corresponds to state *A* (the precise nature of this correspondence being the point at issue); and

(*ii*) I am in pain, if and only if, my *C*-fibers are stimulated,

where *C*-fibers are some identifiable item in the brain of the speaker, Jones, say.

Putnam examines the standard dualist use of statements like (*ii*), and their claim that (*ii*) cannot be analytic (and so 'be in pain' *means* 'stimulation of *C*-fibers') because the verification criteria of being in pain, on the one hand, and of having one's *C*-fibers stimulated, on the other, are quite different. The dualist contends that being in pain is a direct observation for Jones, in that it makes no sense to ask *how* Jones knows he is in pain, though it makes perfectly good sense to ask how he knows his *C*-fibers are being stimulated.

The heart of Putnam's case is the true statement that, similarly, it makes no sense to ask how the *TM*, that prints out (*i*), knows it is in state *A*. Hence, argues Putnam, the very same dualist arguments must also defend the non-analyticity of (*i*), and hence the distinction between the bodies and minds of *TM*s.

3 ABSTRACT AND REAL MACHINES

I believe that there is a straightforward sleight-of-hand trick here with the notion 'a *TM* is in state *A*', and that, for any real *TM* whatever, it makes perfectly good sense to ask how it knows it is in state *A*, and hence the analogy between the verifications of (*i*) and (*ii*) breaks down immediately.

The power of Putnam's argument lies in discussing only the abstract machine when introducing the notion of 'state', and then, when drawing conclusions about real *TM*s, continuing to use 'state' as an undefined abstract term even though, as we shall see, the very process of setting up a real, as opposed to an abstract, *TM*, must have made the notion 'state' concrete too.

In the abstract *TM* 'state' is not defined, nor need it be. As Putnam puts it: '. . . the "logical description" of a *TM* does not include any specification of the physical nature of these "states" . . . a given *TM* is an *abstract* machine'.[1]

The states are just primitive items in terms of which the *TM* is defined. Thus, items in the machine table are of the form (taking account of the row and column in which the instruction appears): '. . . If the machine is in state *A*, and it reads symbol *s*1 then (1), print symbol *s*2 (after erasing current symbol), (2) shift into state *B*, and (3), scan the square to the immediate right. . .'.

So, as I said earlier, the abstract *TM* is not really a machine at all, but a format for describing a clear notion of 'instruction'. It does not actually carry them out, but relies on the imagination of the reader to see how easily it would be to do so, by reading the table, moving the tape *etc.*

But as soon as the *TM* becomes concrete, a different situation arises, even if the concreteness is provided only by pieces of cardboard and a reading and moving and writing person, who carries out the above instructions by hand in a 'mechanical' way (the nature of the person, or that it *is* a person, is of no importance here).

Consider such a pencil and card machine in which the current state is indicated by a pack of cards with upper case letters on them, *A*, *B*, *etc.* so that the state at any time is indicated by the card face uppermost on top of the pack. Notice that the current state must be indicated somehow: it cannot remain an undefined notion or we cannot proceed with the game.

Nothing could be simpler than this toy, but real, *TM*. When the person reads the machine table item above, 'if the machine is in state *A*, and . . .' he looks at the top of the pack to see if an *A* is showing, if it is he goes on, if not he looks at the next machine table item, and acts in an appropriate way.

It is a perfectly clear question 'how does this *TM* know it is in state *A*?' and the answer is 'by looking at the top card to see if *A* is there'. Whenever the *TM* becomes real, some such definition of 'state *A*' must be given, so that 'state *A*' then becomes a derivative notion. It may remain in what Putnam calls the 'logicians' description' but only by courtesy, as it were, for it has no function now that a real *TM* has been defined. This is the crux of the matter.

---

[1] Putnam [1964], p. 147.

Conversely, if no such definition, or procedure, has been given, then we still have only the abstract *TM* and the real *TM* cannot run. But, once the definition or procedure has been given, then the key 'how' question has a clear answer, and Putnam's argument from analogy collapses.

Moreover, Putnam's analogy will not hold if he does not define 'state A' and is left with only the abstract *TM*. For the analogy will then necessarily be with a disembodied, or abstract, person without *C*-fibers at all, and no results about the brain or real persons could follow. If, as Putnam says, the question 'how does the [abstract] computer know it is in state *A*' is silly, it is the category silliness of 'how does the Average Man tie up his shoe laces'.

In sum, 'state *A*' is undefined in the abstract *TM*, and rightly so. Hence any question as to how the abstract *TM* knows it is in state *A* is, as Putnam says, absurd. But in any real *TM* the question is perfectly sensible. Putnam confuses these two situations throughout by continuing to speak of an elusive undefined 'state *A*' with respect to a real *TM*, so that in his discussion of (*i*) above, he simply will not allow that to know flip flop 36 is on, *is* to know the machine is in state *A*. Yet since the machine is real and not abstract, we can just demand that he tell us what it is to be in state *A*, and once told, his case fails, for state *A* then has no unexhausted content.

This can be seen at its clearest when Putnam considers the possibility that, since (*i*) is said to be a 'synthetic proposition' for the machine, '. . . the machine might be in state *A* and its sense organs might report that flip flop 36 was *not* on. In such a case the machine would have to make a methodological "choice"—namely, to give up (*i*) or to conclude that it had made an "observational error" . . .'[1]

But the error is Putnam's: talk of flip flop 36 shows we are talking of a real *TM*, and so the 'deciding it is in state A' has no sense here apart from whatever procedure or definition has already been chosen to exhaust the notion 'state A', and no such problem can arise. If such an assignment has not been made, the machine is still abstract and talk of flip flop 36 is out of place.

Putnam's claim[2] that 'I am in state *A*' and 'flip flop 36 is on' are 'clearly non-synonymous in the machine's language by any test . . . the machine has to use different "methods of verification" . . . .' is simply and straightforwardly false for any real *TM*.

4 PUTNAM'S COUNTER ATTACK

Putnam anticipates to some extent this attack on his use of the abstract *TM*. He correctly notes

---

[1] Putnam [1964], p. 139.

[2] Putnam [1964] p. 153.

> As soon as a *TM* is physically realised, however, something interesting happens. Although the machine has, from the logicians' point of view, only the states *A*, *B*, *C etc.*, it has from the engineer's point of view an almost infinite number of 'additional states' . . . we shall call these 'structural states'. . . . if the machine consists of vacuum tubes, one of the things that may happen is that one of its vacuum tubes may fail—this puts the machine in what is from the physicists if not the logician's point of view a different 'state'.[1]

But the false move is still there: the retention of an 'uninterpreted state' in the real (vacuum tube) *TM*. To see this clearly, we ask how can Putnam know that the failure of the vacuum tube will *not* cause the machine to go into a different logical state? That is, how can Putnam's logician know that the failed tube is not part of the procedure that defines, say, state *A*?

He can know that only if either (*a*) he knows that the logical states have nothing to do with the 'structural states', and so no tube's failing matters in the least. In that case the states belong only to the abstract *TM*, and no results about real machines or persons can follow. Or, (*b*) he has detailed information and knows that the particular tube that failed will not cause the machine to change state, since it is not, as it were, mentioned in the necessary conditions for being in its present state. In that case we simply shift the discussion to another failing tube where Putnam's point will not hold, and one such tube is sufficient to refute him. If there is no such tube then we have consequence (*a*) above.

5 CLARKE'S ARGUMENTS

Clarke's principal argument is that '. . . there is an indefinite number of possible self-consistent, but mutually exclusive, logical descriptions of a *TM* which is in any given physical state or going through any given physical process.'[2] It is important to note that Clarke refers to 'physical state' and 'physical process' here, and that he is not allowing Putnam the merely abstract *TM* with its uninterpreted states. Clarke makes his point in detail with regard to a conventional general purpose computer, and identifies what Putnam calls the 'logical description' (in terms of states) with the machine's program. We shall have reason to question this identification later, but for the moment let us pass to Clarke's question: '. . . on what grounds can it be claimed that there is an indefinite number of possible true logical descriptions of a *TM* which is passing through any given set of physical states?'[3]

Clarke gives what we may call the 'code cracker's answer' that, if we examine the machine's 'encoded output messages', there will be an indefinite number of possible self-consistent translations of these. This argument is extended (p. 4) to any attempt to infer the program, or machine table, from the 'physico-chemical constitution' of the machine, which would, presumably, mean examining the content of its registers. Later (p. 5), Clarke argues that Martians observing a computer's input and output could not decide 'what it was doing', in the sense of whether it was composing music or playing chess.

---

[1] Putnam [1964], p. 147.

[2] Clarke [1972], p. 2.

[3] *Ibid.*, p. 3.

Later still (p. 6), Clarke entertains the possibility that the machine's program may in fact be performing two such tasks simultaneously 'in the sense that identically the same physical operations are interpretable as performing two quite distinct tasks'. Clarke's purpose in all this is to claim that the last possibility does not arise for persons, and hence that Putnam's machine-person analogy breaks down at that point. '... at some level we are thinking what we are thinking, regardless of any interpretations that may be adduced' (p. 9).

Machines are quite different says Clarke, for '... as we have seen an indefinite number of quite distinct logical descriptions can be given of one and the same set of physical process. What the machine is "doing" in a logical sense is in certain given circumstances simply a matter of how we choose to interpret its activities' (p. 9).

I do not want to discuss Clarke's points about the nature of human thought in any detail. A brief rummage through novels, psychoanalytic experience and one's own interior life should produce an indefinite number of examples with the required ambivalence, and, to maintain his position, Clarke would very soon be reduced to declaring by fiat that the examples produced of human thought have only one 'real' interpretation. Furthermore, it is easy to describe combined human performances and introspections that approximate, as closely as necessary, to Clarke's description of machines. For example, a father, calculating his household accounts while, at the same time demonstrating to his son the principles of arithmetic, may be genuinely at a loss to answer the question 'but what are you *really* doing?'.

However, it must be admitted that Clarke has anticipated discursive, anecdotal rebuttals along the lines of the above, by himself pointing out (pp. 7–9) that a person's thoughts 'can be described from various aspects and at various levels of generality'. He also raised the question of the unconscious content of thinking. In doing so Clarke has himself weakened his denial of the person-machine analogy in the form in which he had stated it. My own substantial points concern Clarke's arguments about machines, and I shall now turn to them.

6 CLARKE'S ARGUMENTS ABOUT MACHINES ARE MISCONCEIVED

Clarke's point that the program (or 'what the machine is really doing') cannot be definitively and reliably inferred from low level information about the machine's processes is a correct and important one. The point should, I think, be put in terms of the examination of the machine's registers or machine code, rather than in terms of output and input, since most modern computers capable of the tasks Clarke postulates have input and output somewhat closer to English than the inscrutable punched cards Clarke describes. However, the adjusted point remains important. Its function, I believe, is to defend the privacy of machines, which Clarke would not want to do, though doing that would, *a fortiori*, defend the privacy of human minds (from decoding by means of brain pulse examination, say) a task that Clarke would almost certainly find congenial.

However, Clarke's version of the point contains an ambiguity that would become increasingly important, were we to pursue this line of argument. It is unclear from his several versions whether he is claiming 'There are always a number of possible self-consistent interpretations of a computer's activity, therefore (1) there is no one thing it can be said to be really doing, or (2) we cannot know *which* one thing it is really doing'.

(2) may well be true, if the non-derivability of the program from the register changes is allowed, but (1) is almost certainly false. What an adequately working computer is 'really doing' is what it was programmed to do and that is invariably one activity. This point becomes quite clear when we remember that the program for complex activities of the sort Clarke describes is normally expressed to a machine in a 'high level' programming language. Such languages have names LISP, FORTRAN, and PL-1 *etc.* and are the form in which the programmer conveys his program to the machine which then, before carrying out the instructions, 'translates them down' into a machine language, which mentions the actual registers or switches of the machine. The high level instructions normally translate into many such machine instructions. It is possible to program in machine language, but as a matter of practical fact, very difficult to do for complicated tasks—it is rather like reading a novel one letter at a time.

These high level languages may approximate to natural languages, and it is the program at this level that represents what the programmer wants done and therefore what the machine is really doing. Clarke will be in difficult straits if he argues that this high-level program can always be interpreted in some other way. If it can, then, in the same sense, a speaker of an English sentence can always be interpreted as saying something other than what he was saying, as it were. That may indeed be true, in some sophisticated sense, but I am not sure Clarke really wants to be committed to it. As a thesis about ordinary usage, it might well turn out to have disastrous consequences for some of his other arguments based on usage, such as his argument about the essentially unambiguous nature of human thought.

Readers who remember the last section will realise that Clarke's arguments have now been stood on their heads, for, using considerations like Clarke's, I have argued that complex computers are always really doing one thing, whatever the possibilities of additional rebarbative interpretations of their machine code, which may leave us unable to infer that one thing from the machine code.

7 CLARKE AND PUTNAM

Clarke's arguments simply do not tell directly against Putnam's analogy and dependent thesis. The reason is that Clarke does not force Putnam to declare whether he is describing abstract or real *TM*s. Clarke is clear that his own arguments concern real *TM*s: indeed, it is not clear that his arguments would tell at all against abstract *TM*s since it could be argued that all reinterpretations of a given *abstract TM* are trivial redescriptions of the same machine table.

Thus, in describing real *TM*s Clarke identifies Putnam's 'logical description' with a machine's program. But Putnam need not accept this identification at all, and indeed the argument of Putnam's that I considered in detail suggested that he would not, for Putnam's logical description is in terms of 'states' and if a program is interpretable, as it must, be, as a sequence of the physical states of the machine, then this identification will define 'states' procedurally in a way that, as we have seen, Putnam is unwilling to do. For Putnam's whole argument assumes that 'state' has not been defined without residue in that way.

Thus, Clarke's central argument tells, if it does tell, only against an analogy between persons and concrete *TM*s. As I showed in section 3 Putnam is talking about abstract *TM*s, and *his* central argument does not hold for concrete *TM*s. Thus, on the face of it, Clarke's argument does not connect with Putnam's. What Clarke should have done, in my view, was to show that only concrete *TM*s were of interest in any analogy with persons, and that Putnam ought to have been talking about them.

Yet, as I showed in the last section, even if Clarke's arguments could be brought to bear on Putnam the former's claims about the relative ambivalence of human and machine processes are equally plausible the other way around. Indeed, we might well choose to show more interest in inscrutable ambiguous machines than in transparent single-minded persons.

8 PRIVACY AND 'TURING'S TEST'

I think a general point of interest emerges from Putnam and Clarke's arguments about the possible privacy of a machine, or what Putnam would call, satirically, its infallible knowledge of its own state. Both Clarke and Putnam argue for this notion in some sense. Putnam does so ironically: 'Anyone who wishes, then, to argue on this basis for the existence of the soul will have to be prepared to hug the souls of *TM*s to his philosophical bosom!'[1]

If persons have privacy, says Putnam, then so do *TM*s and, conversely, if *TM*s do not, then neither do persons. And it is clear that it is the contraposed form that he has in mind.

Clarke on the other hand, uses the 'code cracking argument' which, I would claim, does establish the privacy of machines in another sense. Clarke does not seem to notice this, and has no desire whatever to establish anything other than the *ambivalent* privacy of machines, which he evidently does not consider privacy at all. What is clear in all this is that Clarke is for the privacy of persons and Putnam is against it. Neither of them considers the privacy of machines *seriously* though, as I said, Clarke ought to because its establishment establishes that of humans *a fortiori*.

I think there is an interesting point here, in that the true privacy of machines comes not downwards, as it were, from Putnam's uninterpretable notion of state, but upwards from our inability to infer the machine's real activity determinately from the content of registers. Anyone who argues that an experienced programmer can detect the 'sort of general shape' of high level language *X* in large blocks of machine code, ignores the possibility that the code may have been written in language *Y* with the syntactic style of *X* to create just this confusion. (Just as one can murmur English with a German intonation and lead a distant listener to believe he is listening to unintelligible German).

What I believe to be of interest in the notion of 'machine privacy' arises in connection with the possibility of an adequate simulation of a person by a machine. This possibility is usually discussed in connection with the so-called Turing test[2] which has no connection, other than Turing's name, with the *TM* discussed earlier.

Turing suggested that if a machine could sustain a dialogue with someone adequately for a sufficient period of time, and in such a way that the person was unable to decide whether he was talking to a person or a machine, then we should cut through the conventional conceptual puzzles about whether or not a machine can think and simply admit that such a machine *was* thinking.

[1] Putnam [1964], p. 153.

[2] Turing [1950].

I would suggest that one component of a truly convincing simulation would be a display of the sort of machine privacy discussed by Putnam and Clarke. That would mean that we, as observers, agreed to, or were persuaded to, ascribe to the machine the final authority as to what state it was in, in the way that we now do for persons. For, if Jones says he is in pain then the neurosurgeon exploring his brain with a probe has, at some point, to accept what Jones says, regardless of what his theory about Jones's brain leads him to expect.

So that, if a machine printed out that its memory was suffering some disability, then we might well be persuaded by past experience with such machines to continue examining its hardware for faults, even though none was found in the part of its workings where we would expect it. These 'states', about which the machine might know better than us, are a more direct analogy to Jones's pain than Putnam's uninterpreted and elusive notion of *TM* state. The addition I am suggesting to the Turing test would turn it into something rather different from the original, but we can imagine a dialogue with an unseen entity that discussed its mental states and their relation to what skilled diagnosticians were able to find about its physical constitution. We have, I think, some fairly clear notion of what constitutes the sort of authoritative statement that persons make about their states, and we should expect no less from any machine that we were going to pass on the Turing test. It might well be possible to give the relevant computer/brain structural data simultaneously to the human participant in the dialogue without violating the conditions that Turing set up for the test, for the person might be told that what he was being shown during the dialogue was the interpreted structural data from the other party, but without being told whether the data was from a brain or a computer.

Putnam also explores the possibility that advances in neurology might give rise to a general linguistic change in which statements like 'mental states are micro-states of the brain' appeared non-deviant through force of usage, in the sense in which Black and others have claimed that such statements are deviant in present speech, because of dualist considerations like those raised at the beginning of this paper.

The weakness in Putnam's suggestion is that this example sentence is so theory-laden that it is hard to imagine a domestic, albeit futuristic, context in which it could ever be at home. My point however is that, on the contrary, the machine privacy and authority I have frivolously imagined is far more likely to give rise to future linguistic shift than the situation (of the loss of human privacy) described by Putnam. In the fairly near future machines will write to us (probably long before they can *say* it) 'I dont care if you cannot find the fault, I tell you my memories about *X* are going'. I find that a far more plausible candidate for future non-deviance, than that Jones saying, acceptably to all, 'I see my monosynaptic excitation is up to X, so I must be in pain/love'. And the force of these two examples is exactly opposite in relation to Putnam's views on body and mind.

The addition of this privacy clause to the Turing test would deal an additional blow to the naive mechanists whose hopes of explaining away the mind-brain problem still lie in the construction of intelligent automata. For such an automaton, having passed this augmented Turing test with flying colors, could no longer be opened (or its program displayed) by the triumphant mechanist with a cry of 'There, see!'. For the authority of what state it was in would now have passed to the machine itself, and the plans/program would no longer be a fully satisfactory explanation of its behavior, since the machine might be in a position to contradict program-based assertions about its state at any given time.

In other words, in the process of becoming mind-like, the machine would *ipso facto* have become private like ourselves. We would then have come almost full circle in this argument, and defined a sense of machine state of which the machine itself has infallible knowledge. The circle is not complete, back to Putnam's premise, because the sense of state now under discussion has no evident relation to a *TM* 'state', and any decision to grant to the machine such an authority would prevent us, at one and the same time, accepting Putnam's invitation to laugh off the authority granted, and to follow his consequence that since the machine cannot really be in that position neither can we. The difference lies in the fact that the state now under discussion is a substantial rather than a fictional one.

At the end of his paper, Putnam remarks:

> . . . if the mind-body problem is identified with any problem of more than merely conceptual interest . . . then *either* it must be that (*b*) some philosophic argument for mechanism is correct or that (*c*) some dualistic argument does show that both human beings and *TM*s have souls. I leave it to the reader to decide which of the . . . alternatives is at all plausible.

In this last section I have tried to suggest, that by pushing Putnam's considerations about plausible linguistic change a little further, and taking a more realistic notion of state than the one he abstracted from automata theory proper, the answer to his question is not as obvious as he seems to have thought.

Artificial Intelligence Laboratory,
Stanford University

REFERENCES

CLARKE, J. [1972]: 'Turing Machines and the Mind-Body problem', *British Journal for the Philosophy of Science*, *23*, pp. 1–12.

PUTNAM, H. [1961]: 'Minds and Machines'. in S. Hook (*ed.*): *Dimensions of Mind*, pp. 138–64.

PUTNAM, H. [1964]: 'Robots: Machines or Artificially Created Life', *Journal of Philosophy*, 61, pp. 668–91.

PUTNAM, H. [1967]: 'The Mental Life of Some Machines', in H.-N. Castaneda, (*ed.*): *Intentionality, Minds and Perception.*

TURING, A. [1950]: 'Computing Machinery and Intelligence', *Mind*, 59, pp. 1–19.

Brit. J. Phil. Sci. 26 (1975), 227–239 *Printed in Great Britain*

Discussions

PUCCETTI ON MACHINE PATTERN RECOGNITION

1 The use of evidence gleaned from experimental sources to buttress an argument which is, in the last analysis, philosophical is a tactic to be treated with respect. In his recent article on the nature of visual pattern recognition,[1] Puccetti has, it seems to me, failed to observe this respect. I shall argue in the sequel that he has, during the course of a discussion which is largely concerned with the possibility of building machines which can see as well as human beings can, misconstrued the experimental evidence which he cites in support of his case. I shall argue further that, even were this not the case, the essentially philosophical point that Puccetti wants to make would still not follow from the premisses which he lays out.

2 Puccetti's exposition begins by discussing the possibility of building a computer (or, more accurately, a computer program) which would have the same sorts of visual capabilities as humans do. Previous discussions which he mentions had asked the interesting question whether we need to know how human visual machinery works before we can build a machine of equivalent capability or, conversely, whether the very attempt to build such a machine would lead us to valuable insights concerning the constraints which govern the way that natural vision systems (or indeed any vision system at all) might work. It will become clear from the sequel that I incline to the second view, but that is not the point here: all that needs to be noticed is that it is possible to conduct an *a priori* discussion of the nature of visual information processing: *a priori*, that is, in the sense that we can discuss the *computational* basis of vision independently of any particular way that it might be implemented in animals or men. To say this is not, of course, to deny the (very likely true) conjecture that when we do finally understand these computational properties, it will turn out that the experimental studies of animal systems and the design studies of artificial systems will converge onto some common mechanisms. However, Puccetti, on the contrary, concludes his preamble rather sharply with the rhetorical question: 'what if human mechanisms do not lend themselves to machine simulation?' The establishment of this pessimistic conclusion I take to be the goal of the argument which he subsequently develops.

Puccetti's approach might be taken, then, to be in some sense 'metaphysical', bearing as it does on what *can* be the case. Such an approach is naturally of great interest if it can be carried through. Since there has been a regrettable tendency in recent philosophy to play down empirical considerations (a tendency which would, no doubt, have caused pain to the great philosopher-scientists like Leibniz, Descartes and Helmholtz), it is very refreshing to see Puccetti's brave attempt to marshal together the considerable body of physiological and psychological evidence which is available concerning, particularly, the lateralisation of intellectual functions between the two cerebral hemispheres, and the perception

[1] Puccetti [1974].

of chess games, and to draw from it conclusions about whether or what machines may eventually be able to see.

3 I shall now examine these two bodies of evidence in some detail, in order to see how they contribute to the pessimistic conclusion. But, first, one must mention the important distinction, which Puccetti mentions, between two paradigms of visual perception, which have traditionally been called *template matching* and *feature analysis*. The first of these models claims that the central process of pattern recognition is that of reducing the proximal retinal stimulus *as a whole* to some canonical form (for example, by some combination of size and orientation normalisation); this canonical stimulus is then compared with a stored model in some straight-forward way. The inadequacies of this paradigm are too well known to need discussion here.[1] The feature-analytic paradigm which is currently favoured claims that visual analysis consists in the abstraction from the retinal array of various sorts of features, for example bars, edges and corners, which are then combined into some representational structure (of a graph-like form, for instance) which may be compared either piece by piece or wholistically with stored representations. The theoretical advantages of this model are manifold: they include economy, computational feasibility and flexibility and, most important, a stress on the *descriptive* rather than *recognitive* aspects of vision. That is to say, it suggests that the major, and computationally most difficult, part of visual analysis is the abstraction of some suitable description of what is seen, rather than some process of straightforward recognition. After all, most of our visual experience is with things that we have not in fact seen before!

Curiously enough, Puccetti attacks this feature-analytic model, claiming that it suffices only to account for what he calls the verbal-analytic-symbolic aspects of vision (which he ascribes to the left hemisphere in man). It seems to me that he should find the stress on the descriptive process attractive to his point of view, for in it wholistic, global processes find a natural place. His taste is for the Gestalt qualities of perception; this would seem to be naturally accounted for in a model which stresses the abstract, rather than immediate, properties of visual data. Perhaps the mistake here is to confuse the modern feature-analytic view with the superficially similar Associationistic views which were popular several decades ago, and which were finally overturned by the Gestalt school. Equally, one must not forget the well-documented weaknesses in a traditional Gestaltist position: weaknesses which do not seem to occupy Puccetti's attention too closely.

Rather curiously also, Puccetti seems sceptical that the feature-analytic mechanisms whose existence in animals has been so convincingly demonstrated have any significance in human visual behaviour. This scepticism is entirely ill-founded. Weinstein *et al.* have demonstrated that human retinal ganglion cells show the same kinds of centre-surround organisation as do the equivalent cells in non-human primates.[2] Still more to the point, Marg and his colleagues[3] have, in an elegant series of studies, shown in awake human subjects that micro-electrodes inserted into the occipital cortex detect the same kinds of stimulus-specific behaviour (especially specific for disks and bars) that others have found in cells of monkey occipital cortex. More indirectly, there is a whole corpus of psychophysical evidence which supports the conclusion that the feature-analytic mechanisms found in higher primates are also active in human perception. In particular, studies of the bandwidth characteristics of spatial-frequency specific channels in human vision may be interpreted in a way which is compatible with the existence of edge-detection mechanisms.[1] It is also interesting to note that the human visual system seems to be responsive to early visual experience in the same way as that of other animals.[2]

---

[1] But see for example Binford and Thomas [1974].

[2] Weinstein *et al.* [1971].

[3] See for example Marg and Rutkin [1968].

I now turn to discuss the central point of Puccetti's exposition: his comments on the lateralisation of visual function between the two cerebral hemispheres. He is concerned to show, on the basis especially of patients who have undergone section of the commissures connecting the left and right halves of the forebrain, that the left hemisphere is responsible for 'verbal' and 'symbolic' behaviour characterised, one might say, by an 'analytic' mode of reasoning. To the right hemisphere he ascribes synthetic and perceptual capacities. From this dichotomy of function he concludes that 'a computer, if it is a computer, cannot operate on other than analytic principles. So it appears the basic *modus operandi* of computers, and hence all future computers as well, is unsuited to recognition'.[3] I shall leave comment on this rather sweeping conclusion to the next section, and turn now to the possible neurophysiological basis for his claim of functional dichotomy.

The problem of the apparent functional asymmetry of the human brain involves some of the very deepest conceptual questions concerning its operation, and so must be approached with caution. On the basis both of the analysis of the deficits caused by unilateral lesions, and of the elegant series of studies by Sperry and his colleagues on split-brain patients, one might easily come (as Puccetti does) to the overly simplistic conclusion that there is a clear-cut division of labour between the left and right. That this conclusion *is* overly simplistic is suggested by studies on the selective interference by one hemisphere in the functioning of the other. For example, reaction time to tactile stimulation of the finger tips is impaired by a task involving acoustic memory for letters, and *vice versa*.[4] In a shadowing task, Allport *et al.*[5] showed that errors of timing occur in playing the piano when speech is being shadowed, suggesting at the very least that speech and musical tasks cannot be separated (and surely music is a prime candidate for a synthetic, Gestalt activity, rather than an analytic one?). In another study, Broadbent and Gregory[6] showed that efficient reaction to one side of the visual field using the corresponding hand (*i.e.* the hand which is ipsilateral to the destination of optic fibres from that half-field) interfered with efficient reaction on the other side of the visual field using the other hand. Now if it is really the case that certain functions are completely lateralised, it surely ought to be possible to find combinations of tasks which can be performed without mutual interference. This has proved to be very difficult. Professor Berlucchi[7] has shown that reaction time for recognising letters is shorter in the right half-field (*i.e.* the left hemisphere), while reaction time for facial recognition is shorter in the left half-field (*i.e.* the right hemisphere). This seems, on the face of it, to support

---

[1] See for example Klein and Stromeyer [1974].

[2] Freeman and Thibos [1973].

[3] *Op. cit.* p. 145.

[4] Broadbent and Gregory [1965].

[5] Allport *et al.* [1972]

[6] Broadbent and Gregory [1967].

[7] Berlucchi [1974].

Puccetti's point of view, but then Yin[1] has demonstrated that patients with trauma of the *right* occipital region, while showing a handicap in recognising snapshots of faces presented the right way up, were *less* handicapped than healthy subjects when the photographs were presented upside down. This suggests that the recognition of faces is a highly specialised task whose implications for general pattern recognition should be viewed with some scepticism. The consensus of workers in the field at the moment seems to be that any clear-cut statement of functional dichotomy is likely to be misleading, and that much more attention needs to be given to the phylo- and onto-genetic factors which generate the predispositions that do seem to be observed.

Having established that the division of labour which Puccetti advocates is, to say the least, not proven, it is now time to consider the other limb of his argument, *viz.* the evidence concerning human perception of complex situations, and of chess games in particular. It is well-known that chess-masters, while no better than anyone else at traditional memory tasks, are remarkably good at extracting, during short exposures, the crucial aspects of the arrangment of chess-pieces on a board. Not surprisingly perhaps, they excel in this task only when the configuration is *meaningful* in chess terms: their performance is no better than average when the pieces are disposed *at random*. From this insight, Puccetti draws the conclusion that 'what would appear crucial to master level play is not superior *general* knowledge of chess possibilities and probabilities, but a way of organising the visual information into configurations that lay bare the dynamic interrelationships of the pieces. And this is a Gestalt-synthetic process, not a sequential, analytic one'.[2] This inference is absolutely right, but surely uncontroversial. Neither is it restricted to chess. Modern psychology (especially the developmental schools represented by such workers as J. Piaget and J. Bruner) has been much concerned with exploring the ways that movement and perception interact, and from there it is a small, albeit difficult, step to considering other kinds of dynamical relationships.

4 This last point has not been lost on workers in artificial intelligence (*AI*); indeed a large part of what has come to be known as the 'representation problem' in *AI* is concerned precisely with the question of what sorts of representational formalism are adequate to capture the 'non-visual' information which is embodied in visual scenes. This question is being vigorously pursued in *AI*, and many interesting results have already been obtained.[3] Furthermore, some considerable success has been achieved in the building of artificial vision systems[4]; this effort has certainly not ignored (and *could* not ignore) considerations of the sort which Puccetti wants to bring to our attention.

Of course, *AI* is proceeding on the assumption, which Puccetti denies, that computers *are* adequate vehicles for expressing representational formalisms, and for testing out their efficacy. It seems to me to be sheer obscuratism to make claims about what computers can and cannot do; in any case, this controversy has been well worked out in the discussions which followed the publication of

---

[1] Yin [1970].

[2] *Op. cit.*, p. 150, italics in the original.

[3] The literature is huge, but for a representative sample of the different viewpoints, their advantages and weaknesses, see Hayes and McCarthy [1969] and Winograd [1974].

[4] For a review, see Binford and Thomas [1974].

Hubert Dreyfus's book on the subject.[1] Neither is there any particular value in thinking of computers as representing some 'analytic' as opposed to 'synthetic' mode of reasoning, unless some much clearer distinction is made between these two modes than Puccetti for one seems prepared to make. Computers as such are not committed to some particular form of reasoning—only the people who write programs for them are. There is certainly no reason in principle (or even in practice) why the 'Gestalt-synthetic' point of view should not be imbedded in a program: the only requirement is that the algorithmic properties of this point of view should be adequately and coherently expressed. To require that is surely no more than to require that an advocate of that view express his ideas in some cogent manner which meets the normal canons of scientific exposition. To make claims about what computers will eventually be able or unable to do is an even greater exercise in futility. Puccetti quotes another sceptic, who states 'the interesting fact that comes through . . . is that we can perform all sorts of intricate operations that we really don't understand . . . we cannot dissect the activity in a step-by-step manner. This is the principal reason why the computer is limited in its abilities at present. Furthermore, it is not clear that the situation will ever significantly improve'.[2] I trust that my precis of this quotation does it no injustice: in any case, Bellman seems profoundly pessimistic. His pessimism seems ill-founded, or at least founded only on the realisation that we don't know everything there is to know about our intellectual capacities and mechanisms. One needs more than pessimism to pass from the current limitations on our knowledge about brains to a sweeping generalisation about limitations on our future knowledge. One needs a firm metaphysical argument of the sort which seems, on close examination, to be missing from Puccetti's article.

ARTHUR J. THOMAS
Artificial Intelligence Laboratory,
Stanford University

REFERENCES

ALLPORT, D. A., ANTONIS, B. and REYNOLDS, P. [1972]: 'On the Division of Attention: A Disproof of the Single Channel Hypothesis', *Quarterly Journal of Experimental Psychology*, **24**, pp. 225–35.

BARBER, F. H., HOBSON, R. R. and WEINSTEIN, G. W. [1971]: 'Extracellular Recordings from Human Retinal Ganglion Cells', *Science*, **171**, pp. 1021–2.

BELLMAN, R. [1968]: *Vistas of Modern Mathematics*.

BERLUCCHI, G. [1974]: 'Cerebral Dominance and Interhemispheric Communication in Normal Man', in F. O. Schmitt and F. G. Worden (*eds*.): *The Neurosciences: Third Study Program*. Cambridge, Massachusetts: The MIT Press.

BINFORD, T. O. and THOMAS, A. J. [1974]: 'Information Processing Analysis of Visual Perception: A Review', *Stanford A.I. Memo*, **227**.

BROADBENT, D. E. and GREGORY, M. [1965]: 'On the Interaction of *S-R* Compatibility with Other Variables Affecting Reaction Time', *British Journal of Psychology*, **56**, pp. 61–7.

BROADBENT, D. E. and GREGORY, M. [1967]: 'Psychological Refractory Period and the Length of Time Required to Make a Decision', *Proceedings of the Royal Society of London* (*Biology*), **168**, pp. 181–93.

DREYFUS, H. [1973]: *What Computers Can't Do*. New York: Harper and Row.

---

[1] Dreyfus [1974]. This has been extensively reviewed, *e.g.* Williams [1974], Pylyshyn [1975] and Wilks [1975].

[2] Bellman [1968], p. 109.

FREEMAN, R. D. and THIBOS, L. N. [1973]: 'Electrophysiological Evidence that Abnormal Early Visual Experience can Modify the Human Brain', *Science*, **180**, pp. 876–8.

HAYES, P. and MCCARTHY, J. [1969]: 'Some Philosophical Problems from the Viewpoint of Artificial Intelligence', in D. Michie (*ed.*): *Machine Intelligence*, **4**. New York: American Elsevier.

KLEIN, S. and STROMEYER, C. F. III [1974]: 'Spatial Frequency Channels in Human Vision as Asymmetric (edge) mechanisms', *Vision Research*, **14**, pp. 1409–20.

MARG, E. and RUTKIN, B. [1968]: 'Receptive Fields of Cells in the Human Visual Cortex', *Experientia*, **24**, pp. 313–6.

PUCCETTI, R. [1974]: 'Pattern Recognition in Computers and the Human Brain', *British Journal for the Philosophy of Science*, **25**, pp. 137–54.

PYLYSHYN, Z. [1975]: 'Minds, Machines and Phenomenology', *forthcoming* in *Cognition*.

WILKS, Y. [1975]: 'Dreyfus' Disproofs', *forthcoming*.

WILLIAMS, B. [1974]: 'Review of H. Dreyfus "What Computers Can't Do" ', *The New York Review of Books*, November.

WINOGRAD, T. [1974]: 'Five Lectures on Artificial Intelligence', *Stanford A.I. Memo*, **246**.

YIN, R. K. [1970]: 'Face Recognition by Brain-Injured Patients: a Dissociable Disability?', *Neuropsychologia*, **8**, pp. 395–402.

MACKIE ON DISPOSITIONAL PROPERTIES

In chapter IV, section 2 of his *Truth, Probability, and Paradox* J. L. Mackie attempts to show that dispositional properties do not form a 'special ontological category' and that no properties are dispositional 'in themselves'. He argues that where a body is ordinarily described as having a certain dispositional property, it will have some non-dispositional property which accounts for the body's possession of the apparently dispositional property, and that the property to which the dispositional style of description applies is in fact this underlying non-dispositional property. This claim has of course been made by other writers, notably by D. M. Armstrong in his *A Materialist Theory of the Mind.*[1] Mackie however does not accept the *a priori* argument proposed there by Armstrong in defence of this position. Mackie's own argument relies on showing that, if we suppose this identity to hold, much becomes comprehensible that has to be left unexplained on the alternative view, according to which there are irreducibly dispositional properties, and that indeed this alternative view commits its holders to accepting some highly implausible propositions. I wish to point out that Mackie's main argument in favour of his position is inconclusive, and therefore unsuccessful. In his exposition Mackie's objections to the alternative view are intertwined with the recommendations of his own view, and in order to see what he considers the main merit of his position, it will be helpful to consider also one of the main objections that he makes to the alternative view.

Against irreducibly dispositional properties Mackie argues that 'it seems quite unnecessary to postulate them. Why should we insert this extra element between the non-dispositional basis and the causal behaviour?' (p. 137). It is generally accepted, Mackie points out, that an occurrence that is regarded as a manifestation of a physical dispositional property can be causally accounted for by reference to some such 'basis' as the molecular structure and molecular behaviour of the body, which, together with the occurrence of some appropriate event, will bring about the disposition-manifesting occurrence. No reference is made to a distinct dispositional property in such a causal explanation of the manifestation of the dispositional property. The supposedly distinct dispositional property thus appears to be redundant.

---

[1] Armstrong [1968], chapter 6, section VI.

Mackie suggests that the real distinction is not between ontologically distinct dispositional and non-dispositional properties, but between dispositional and non-dispositional styles of description of properties that are themselves non-dispositional, or, again, between knowing a property as a dispositional property and knowing it as a non-dispositional property. Where we do not know what the non-dispositional basis is, we shall have to use the dispositional style of description. 'But this would not make *what is there* dispositional, and it would not give the concept of disposition or power any ontological or metaphysical role' (p. 136).

In order to see clearly the deficiency in Mackie's position it will be helpful to take one of his illustrations, which may fairly be taken as a paradigm for the relation between the dispositional property and its non-dispositional basis, as Mackie conceives the relationship. A piece of iron will become brittle at very low temperatures.

> We might describe the difference between an iron bar at a very low temperature and the same bar at an ordinary temperature by saying—if only we knew this—how the molecules were arranged and moving. Or we might describe this difference by saying that at the low temperature the bar is such that if it is bent, it breaks. The latter is a dispositional description, the former a non-dispositional one. The latter description points to the same difference that is explicitly described by the former description . . . (p. 136).

We may notice first that what is identified with the dispositional property of the body, brittleness, is not itself a property, either of the body or of its constituent molecules. It is one or more states of the body: the moving in a certain way of the molecules, and their being arranged in a certain way. Although philosophers tend to interpret the notion of a property in a rather broad way, compared to the everyday, or scientists' everyday, use of 'property', it would be unusual even for a philosopher explicitly to represent the fact that a body was moving in a certain way as its having a certain property. Mackie might however argue that the important point is that the dispositional property of the body can be identified with *something* non-dispositional; whether the identification is with another property or with a state or states, is not material. I shall not pursue here any reservations that may be felt about the identification of a property with something that is not a property, which is common in the discussion of this particular issue, although not, I think, in other contexts. For even if this identification is accepted, Mackie is open to a fundamental objection.

In his example Mackie mentions the way molecules are moving and the way they are arranged as the non-dispositional features of the situation that correspond to the dispositional property of brittleness. However, we may ask why they are moving in those ways and why a certain arrangement is preserved amongst them. A full answer to these questions will include some mention of the properties of the molecules. Again, if we ask why they behave as they do when the iron bar is bent, the full answer will also mention some properties of the molecules.

Although we would ordinarily be satisfied to have the manifestations of brittleness explained in the way Mackie suggests, this is because we shall be taking for granted certain facts about the sorts of things the molecules are, that they have certain properties and lack others. Parallel points can also be made about molecular structure and atomic structure, which Mackie assumes will feature in the description of the bases of various dispositional properties. In explaining why a thing has a certain molecular structure, the properties of the constituent atoms will be mentioned, and in explaining why an atom has a certain structure, the properties of the sub-atomic particles will be mentioned.

It appears then that in a full account of the occurrences that we regard as manifestations of dispositional properties, the *properties* of constituents will feature, as well as the non-dispositional states on which Mackie invites us to focus our attention. With regard to any one of these properties we may ask, Is it a dispositional property or is it non-dispositional? If all such properties were *clearly* non-dispositional, Mackie's claim that there are no irreducibly dispositional properties might seem to have much to recommend it. But this is of course not generally agreed. Suppose then that, as seems probable, some such properties appear to be dispositional. How could it be shown that they are in fact non-dispositional? Clearly Mackie cannot rely on a reapplication of the line of argument that has just been discussed, since a full account of the causes of occurrences manifesting dispositional properties will always include the mention of properties of the constitutents featuring in the behaviour and arrangements that form the 'non-dispositional basis', and an exactly parallel question would arise concerning these properties. In order therefore to show that there are no irreducibly dispositional properties some other arguments must be adduced, either some other general, metaphysical considerations like Mackie's, or else a more pedestrian, empirical demonstration that the small number of properties that will appear in the most fundamental accounts of physical phenomena are by general admission non-dispositional properties.

J. W. ROXBEE COX
University of Lancaster

REFERENCES

ARMSTRONG, D. M. [1968]: *A Materialist Theory of the Mind.*
MACKIE, J. L. [1973]: *Truth, Probability, and Paradox.*

NEWCOMB'S PARADOX*

The following game has been discussed lately by various writers: One has a choice between taking box II only or taking box I and box II, where it is given that box I contains $1,000, and that a predictor has put $1,000,000 in box II if he predicts I will take box II only, and has left box II empty if he predicts I will take both boxes. It is also given that the predictor has played this game thousands of times without losing.

---

* I am grateful to my colleague, W. K. Goosens, for discussions which were helpful in thinking about this topic. Since writing this, I have had the pleasure of reading Isaac Levi's excellent note on this topic, which is forthcoming in *Theory and Decision.*

In the original published discussion of this game, Robert Nozick[1] exhibits considerable familiarity with the techniques of decision theory and with the literature of the subject. But after writing at length, he leaves off with no firm conclusion, and with the indication that this game, which he attributes to William Newcomb, poses a formidable intellectual problem.

Nozick's paper was made the basis of a column in *The Scientific American* by Martin Gardner.[2] Gardner suggests that Newcomb's game leads to a paradox with important implications for the problem of free will. To examine his argument let us call the proposition that the predictor is right, P, the proposition that Box II contains \$1,000,000, M, and the proposition that the player chooses box II only, C. We can then treat 'choose both boxes' as $\sim C$, and 'box II is empty' as $\sim M$.

The problem, then, can be viewed as a decision between C and $\sim C$ when uncertain whether M. Alternately, it can be regarded as a decision between C and $\sim$C when uncertain whether P. There should be no difference in the results, because how the possibilities are partitioned should not matter, so long as they are partitioned exhaustively.

Gardner supposes the probability of P to be .9. Then he reasons that the probability of M given C is .9, and the probability of M given $\sim C$ is .1. Then applying classical decision theory, the expected utility of C exceeds that of $\sim C$. But against this Gardner raises a dominance argument to the effect that $\sim C$ is always the best choice, because taking both boxes must always yield at least as much as taking only one. Thus it is alleged that we have a paradoxical conflict.

This conflict leads Gardner to entertain two conjectures; first, that it is impossible in principle that the probability of a predictor being right in this game should be high, and second, that this fact is evidence for human free will. A similar suggestion has been made by George Schlesinger.

The idea, then, is that if the probability of the predictor being right is high, there is a paradoxical conflict between two decision procedures. A maximal utility argument dictates taking box II only while a dominance argument dictates choosing both. The conflict can only be avoided by denying that the probability of P is high.

This is a fallacious argument. The assumption that the predictor is highly likely to be right does not lead into a conflict between two rational decision procedures. In fact, it does not lead anywhere without further data or, (from a subjectivist point of view), further asssumptions by the chooser as input to the problem.

The crucial question needing determination is this: is the high probability of the predictor's being right independent of the probability of my choosing C or choosing $\sim C$, or is it not?

Suppose it is not. For example, suppose the probability of P relative to C is O, while the probability of P relative to $\sim C$ is 1. The probability of P *simpliciter* can still be high, even 1, if the probability of C itself is sufficiently low. On this account of why the probability of P is high, it is fallacious for Gardner to conclude that since the probability of the predictor being right (P) is ·9, the probability of M given C is ·9. For the probability of P is low relative to C, and is ·9 only because the probability of $\sim C$ is sufficiently high. So in this case, the maximal utility argument is fallacious. But the dominance argument is correct.

---

[1] Nozick [1969].

[2] Gardner [1973].

On the other hand, suppose that the probability of P is high ***independently*** of the probability of C or of $\sim C$. This means that the probability of P relative to C is high *and* the probability of P relative to $\sim C$ is high. Now the maximal utility argument is sound, but the dominance argument is fallacious, for a reason captured in a standard restriction on dominance arguments: a dominance argument for one choice over another does not apply if the choices affect the alternative outcomes. And if P and C are probabilistically independent, then C and M are probabilistically dependent in just the way that violates this restriction.

The above dichotomy, between 'P and C are probabilistically independent' and 'P and C are not probabilistically independent' is complicated by the fact that the dependence could range over a spectrum. It need not be the simple inverse variation I assumed in my example. But I do not think a discussion of this particular complexity is needed for this puzzle.

The probabilistic independence hypothesis could be explained by various theories. One simple one is cheating. Perhaps the predictor is getting his money down by trickery, after you make your choice. This may seem easy to rule out in your own case, but it might not be nearly so easy to rule out as an account of the predictor's past successes. Then if money is your sole object and you do not mind playing along with a hoax, C is the obvious choice.

More exotic theories along this line would include the idea that the predictor is causing money to appear in the right boxes after your choice by real magic, and at the same time altering all records made by any referees, and tampering with the referee's memories as well. This is a logically consistent theory. Whether it would be a reasonable theory would depend on further information. As I stressed earlier, this decision problem does not have, as it stands, sufficient information to determine an answer.

The probabilistic dependence hypothesis could also be explained in various ways. If you are an obsessive who cannot help taking every box you are allowed to take, it's easy to see why a predictor is likely to be right. But perhaps you are sure you are not an obsessive. Perhaps you believe in free will (though I do not find it terribly clear what must be believed to believe that). Let's say you believe you are not predictable, and there just is not information available to anyone at the time the predictor arranged the boxes which would make the probability of either choice high. This does not mean you must give up regarding it as highly probable the predictor will be right. You just need an independence based explanation of why the predictor is likely to be right.

There is one possible source of confusion worth discussing here. Suppose that someone decides that the probability of P really is dependent on C, and thinks that the success of the predictor is based on his having access to deterministic laws. But this chooser figures that these laws are a mystery to him. The truth is that the objective probability of P given X is zero (where X is either C or $\sim C$, you do not know which) and the objective probability of X, on information available to the predictor but not to you, is zero. But all the chooser knows is that the probability of P is high. So now what should he do?

This question touches on a rich source of confusion, which perhaps is why this problem has had such a run. Usually, in a decision problem, the probabilities that count for someone facing the problem as a participant are the ones he thinks obtain, not the ones that do obtain (if there is a difference). Applied to this problem, the point is that it's an empty gesture for the player to say that the 'objective' probabilities of C and P are dependent if he then goes on to reason and act as if he thinks they are independent.

Consider this model, which if carefully restricted, is probably logically consistent, though the question of its reasonableness is another matter—one of those matters left open in the specification of the problem: the predictor has a crystal ball, in which he sees you make your choice. Just before his crystal ball shows the boxes being opened, he makes his disposition on the basis of what he sees you do.

Now it is easy to view this model in terms of probabilistic dependence. The crystal ball's report can be treated as highly reliable, even infallible, evidence available to the predictor relative to which the probability of X (where X is C or X is $\sim C$) is high. So P is high relative to one choice and low for the other. But this representation is not an accurate way of describing the situation relative to your decision problem. For you are treating X as 'whatever I do'. Thus you are treating P and C as probabilistically independent. The determination of probabilistic dependence or independence must be made relative to the way the probabilities are treated in the problem. It may be objected that on the contrary, the problem is, how are the probabilities to be treated? And of course that is a *phase* in assessing the problem, and one for which no basis has been provided, as I have been stressing. But it comes first, logically speaking.

It may be worth noting that the crystal ball model is an example in which the maximal utility argument is applicable and yet there is no obvious infringement on the player's free will. The player is free—he just cannot escape being 'seen' making his free choice. (But of course I am not claiming this case does not involve determinism—that would require settling the free will controversy.)

I do not mean to claim that this crystal ball model cannot rationally be treated in terms of probabilistic dependence. Even this model is under-determined by the evidence. Suppose we simplify the model by removing the crystal ball and having it that the predictor is a simple fellow who just happens to get a pain in his left big toe when a player is going to take both boxes. Suppose that we have a perfect correlation between these pains and player's decisions, and no one can explain why. Now I can *imagine* a case like this which would make *me* certain of probabilistic independence. And this would amount to accepting a probabilistic version of reverse causality.

However, someone else might point out that accepting this would not square well with established science. He might feel he just has to regard the predictor's performance as a hoax, or assume that he is dreaming and hope to wake up soon, or assume that the successes are an extraordinary coincidence.

This absence of a determinate solution is not a serious problem for decision theory. The situation raises a familiar issue in the philosophy of science. If you were to see the sky turn into a golden stair and a choir of angels approaches, announcing the second coming, what should you believe? From a Kantian point of view, it should be impossible for any of us humans to *know* the second coming has taken place, since it represents a gap in 'perception and its empirical advance'. To me, it is imaginable that such a thing should happen and me be there to know it had. But decision theory is under no obligation to find a probability to assign in every fantastic case (or even every actual one). If we connect personal probability judgments with decisions, the theory is waiting to help once those probability judgments are made. But there is no necessary immunity to cases where there is no determinate decision.

If you decide that the predictor is certain to be right independently of your choice *then* it is rationally determined that you should take box II only. If you decide this just cannot be true, and decide he is certain to be right but only depending on your choice, then if you know which choice the chances are for, your decision will be rationally determined again. But you may not be able to decide these questions. And then, you may be in the broad sense, perfectly rational in tossing a coin to decide. What I deny is that there is any case in which there is a conflict in rational decision strategies which can only be avoided by avoiding assigning a high probability to *P*.

George Schlesinger[1] offers an argument which would, if correct, provide an all-purpose defence of the dominance argument (though it may have been intended as a separate argument from the dominance argument): if a sympathetic observer who watched the predictor's move could advise you, he would always advise you to take both boxes, regardless of what the predictor does. But this ignores the fact that the observer may advise this for two different reasons: (1) at least you will get a thousand (2) you might as well get an extra thousand. If the player subscribes to a theory about the predictor according to which whether it is (1) or (2) depends on his choice, the fact that the observer will always advise $\sim C$ may be irrelevant.

It might be replied that once the sympathetic observer checks the boxes, it is too late for the predictor to do anything, so that the fact that the well wisher will necessarily advise $\sim C$ cannot be irrelevant. But this is just to reject the possibility of the sort of probabilistic reverse causality which is implied in accepting probabilistic independence. This may be 'reasonable', but then, given this attitude, there is no conflicting strategy, and it is not necessarily true that this attitude is correct.

I believe that time travel is conceptually incoherent. Some versions of foreknowledge may also be incoherent. But the above 'reverse-causality' hypothesis is not. This is not to say I regard it as scientifically possible. The case is an imaginary one. But in that case, the attitude that it is certain that you will get a thousand if you take both boxes and certain you will get a million if you take box II only is not rationally compatible with the attitude that a sympathetic observer who gets to inspect the boxes after the predictor's move and who will thus recommend taking both boxes, is giving valuable advice.

Perhaps this could be made clearer if we imagine a version in which you, the player, are allowed to be the sympathetic observer. That is, you are allowed to look in the boxes before making your choice. Now the precognition theory about the predictor comes under great strain. But what if he has won many times even with these rules? Well, to keep to the probabilistic independence hypothesis would require supposing that if you 'see' a million in box I, and then choose both boxes, you will find your eyes were deceiving you.

It will not do to reply here that that would be accepting something incredible. For we are being asked to imagine something incredible and are asked how to respond. If you ask a scientist to speculate on what would happen if the law of gravity completely failed, you can hardly complain that his answers fly in the face of scientific theory. But the theory of rational decision is a conceptual theory, which is supposed to work in imaginary cases. And it does, but it is only applicable when the probability judgments required as input are available. If you do not like imagining it possible that a predictor could do these things, then go ahead and imagine it impossible. Then you can regard the predictor's success as a fantastic coincidence, or a hoax, or etc. But in no case will you show that merely assigning a high probability to P leads to a conceptual conflict, because the probability for P by itself, without an answer to the question as to probabilistic dependence, does not lead to a rationally determined choice between C and $\sim C$.

---

[1] Schlesinger [1974].

The most natural cases for the dominance argument are less fantastic. For example, suppose you know the predictor has worked with either superstitious and uneducated people who perceive the predictor as having magic power, or sophisticated people like yourself, and you know that the predictor is otherwise nobody special. His success derives from the fact that the superstitious people respond 'I'm going to do as this marvellous predictor wishes and get a million' while the sophisticated people recognise the profit is all for $\sim C$. In this case, you might say to the predictor 'Well, you know what I'm going to do' and yet no mysterious power is implied.

An important reminder here is that it is part of the description of the problem that you attach no value to crossing up the predictor. If a desire to thwart the predictor is built in, this problem begins to resemble the Surprise Test Paradox.[1] When this is emphasised, the case sketched in the foregoing paragraph is a situation which really is a future tense relative of the prisoner's dilemma. Just as in the prisoner's dilemma it would have paid both prisoners to have been unselfish, so in the case as sketched above, it would pay you to be perceived by the predictor as stupid and superstitious. Instead, you lose a million for being known to be intelligent.

No doubt many interesting cases can be imagined as ways to fill out the sparse details of the puzzle known as 'Newcomb's Paradox'. But I hope I have said enough to bring out that there is no warrant for serious use of the term 'paradox' in connection with the puzzle, and no basis for determining one course of action as the right one without additional information.

JAMES CARGILE
University of Virginia

REFERENCES

CARGILE, J. [1967]: 'The Surprise Test Paradox', *Journal of Philosophy*, **64**, pp. 550–563.

GARDNER, M. [1973]: 'Free Will Revisited with a Mind-Bending Prediction Paradox by William Newcomb', *Scientific American*, **229**, pp. 104–8.

NOZICK, R. [1969]: 'Newcomb's Problem and Two Principles of Choice', in W. Rescher (*ed.*): *Essays in Honour of C. G. Hemple.*

SCHLESINGER, G. [1974]: 'The Unpredictability of Free Choices', *British Journal for the Philosophy of Science*, **25**, pp. 209–21.

[1] For a discussion of this paradox, see my [1967].

Brit. *J. Phil. Sci.* 26 (1975), 241–253 *Printed in Great Britain*

Review Articles

SHOULD PHILOSOPHERS BE ALLOWED TO WRITE HISTORY?*[1]

> '. . . truth will sooner come out from error than from confusion. . . .'
> Francis Bacon, *Novum Organum*, Book II, aphorism XX.

> 'Historical knowledge is the knowledge of what mind has done in the past, and at the same time it is the redoing of this, the perpetuation of past acts in the present. Its object is therefore not a mere object, something outside the mind which knows it; it is an activity of thought, which can be known only in so far as the knowing mind re-enacts it and knows itself as so doing. To the historian, the activities whose history he is studying are not spectacles to be watched, but experiences to be lived through in his own mind; they are objective, or known to him, only because they are also subjective, or activities of his own.'
> R. G. Collingwood, *The Idea of History*, p. 218.

> 'I have tried to describe Faraday, or rather the growth of his personality, in a manner that will defy classification as either civil or intellectual. This is the intended bias of this study. Because of it, I fear you may find reading this volume rather irritating and displeasing. If so, I wish you would try to make light of it and read it as a new kind of historical novel, parallel to today's semidocumentaries, and take my documentation less as a tool of scholarship and more as an attempt to create an air of verisimilitude. This is not to say that knowingly I have forced data or even overlooked discrepancies out of little consideration for truth. Rather, it is to say that I knowingly use my study of Faraday to illustrate a philosophy, and I hope you find this philosophy thought-provoking—perhaps provoking those thoughts that may show me in error and thus enlighten me. For, the main maxim I tried to follow in my writing is, better an interesting error, which may be read and rectified, than a dull truth that will stay put like a heavy tombstone in a forgotten graveyard. I hope I have some measure of success in entertaining my reader.'
> Joseph Agassi, *Faraday as a Natural Philosopher*, p. xiii.

---

* Review of Joseph Agassi [1972]: *Faraday as a Natural Philosopher*. Chicago and London: Chicago University Press. £5·65. Pp. xiv + 359; and of William Berkson [1974]: *Fields of Force*. London: Routledge and Kegan Paul. £6·50. Pp. xiii + 370.

[1] It is well to face at the very beginning an obvious difficulty. I am the author of a large and detailed biography of Faraday and it might be assumed that any review I wrote of rival works would be self-serving. In defence of my acceptance of the task of writing this essay, I should only like to say that I am acutely aware of the problem and have constantly tried to avoid using my privileged position as reviewer to buttress my earlier work. Where I disagree in interpretation with the authors, I have simply referred the reader to my discussion of the subject under scrutiny in order to permit the reader to judge the matter himself. In order to preserve a tone (as well as, I hope, the substance) of objectivity I have cited my work in the third person as Williams [1965], to refer to L. Pearce Williams, *Michael Faraday, A Biography* (London, 1965). As the essay will make clear, an intimate and detailed knowledge of Faraday's life and thought is absolutely necessary for detecting what I feel are the serious, indeed, crippling flaws in the volumes under review. Because my attention is directed at the sources and the intricate arguments of the authors, I have slighted Mr Berkson's work by confining my remarks only to the section he devotes to Faraday. This comprises more than one-third of the volume and I can see no reason why the criticisms aimed at this segment should not, also, have some force in the later parts. But, my knowledge of Maxwell, Hertz and Einstein is not sufficient to permit me to deal with Mr Berkson's discussion of them with the same kind of rigor with which I feel capable of analysing his picture of Faraday.

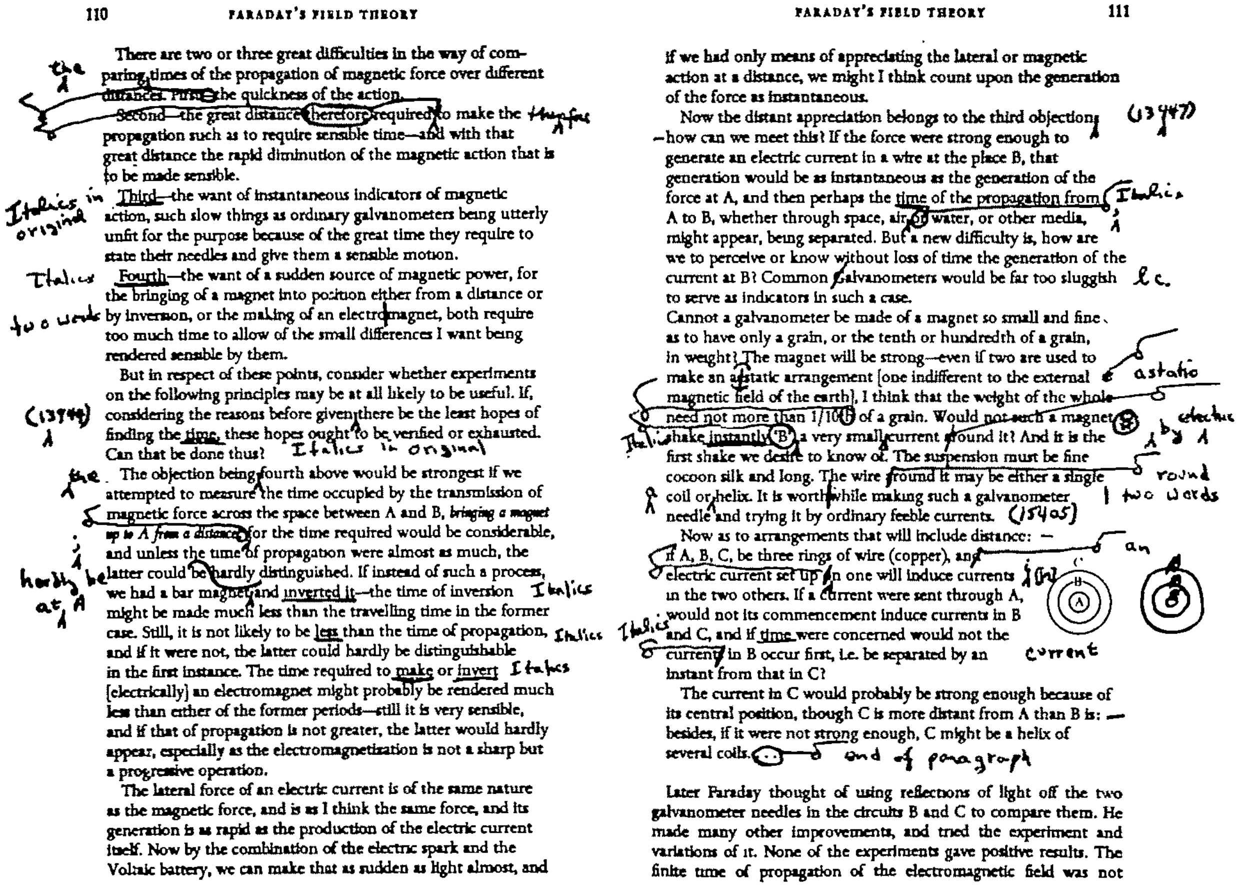
These two works have a great deal in common. Although Agassi concentrates on Faraday, he, like Berkson, has his eye fairly constantly on Einstein and the field theory of the twentieth century. Both volumes, too, are written by scholars whose training has been in philosophy and, indeed, by disciples of the same philosopher, Sir Karl Popper. The treatment of Faraday is very much the same in the two accounts. Both authors see him as marked by his early poverty, struggling against a scientific 'establishment' that paid him or his ideas little heed, and ultimately rejected. The purpose of the two works is to trace his heretical ideas, describe them as a radically new metaphysics, and show how the metaphysics implied an experimental programme that led to mature field theory. Berkson carries the story in detail farther than Agassi, but Agassi says enough to show that he would agree essentially with Berkson's account of later figures. Both Agassi and Berkson overtly make it clear at the beginning that they are using Faraday, in Agassi's words, 'to illustrate a philosophy' and that philosophy is the one associated with Popper's name.[1] This review will be directed at the question, what kind of history is written when it is 'to illustrate a philosophy?'

One of the traits that both authors share that has nothing to do with either history or philosophy is their carelessness in the handling of texts. Without exception, the texts that I compared with the originals were not accurate. Those citations from Faraday's *Experimental Researches* or other writings that Agassi cites on pages 62, 103, 104, 107, 109, 123, 132 and 133 (after which I got tired of checking) all contain errors of transcription. In some cases, this simply consists of italicising words that Faraday has not italicised. Agassi sometimes states, 'Italics in the original' but there are other times (p. 123, for example) when the italics are in the original and he does not so state. It is, therefore, impossible to know whose emphasis is whose and sometimes this is important, for Agassi often sees things in Faraday's writings that appear to have escaped their author's notice. Only in one case is the mis-transcription possibly misleading. On page 109, Agassi quotes Faraday asking about electric curves of force, 'Do they exist also in the electric current wire?' Faraday wrote, 'Do they not exist also in the electric current wire?' The 'not' adds an emphasis that ought to be noted.

Berkson's treatment of his texts is positively scandalous. The accompanying figure (Figure 1) is his version of a section of Faraday's Diary which Berkson states is so important that he must cite it *in extenso*. I have corrected it against the original. Again, most of the sins are venial, but some are not. The sentence marked by an * on the right-hand page is completely garbled. It should read, 'Would not such a magnet shake *instantly* by a very small electric current round it?' The diagram at the bottom is mis-labeled so that the last two paragraphs make no scientific sense. To make matters worse, the footnote refers the reader to paragraph 1342 of the Diary. The passage actually begins with paragraph 13942, some four volumes away.

---

[1] Agassi, p. xiii, as cited in the headnote. See Berkson's Introduction for his discussion of philosophy and science. Berkson there makes clear his debt to Emile Meyerson, as well as to Popper.



110 FARADAY'S FIELD THEORY

There are two or three great difficulties in the way of com-
paring times of the propagation of magnetic force over different the
~~distances. First~~—the quickness of the action.
~~Second—the great distance~~ therefore required to make the thus far
propagation such as to require sensible time—and with that
great distance the rapid diminution of the magnetic action that is
to be made sensible.

Third—the want of instantaneous indicators of magnetic Italics in original
action, such slow things as ordinary galvanometers being utterly
unfit for the purpose because of the great time they require to
state their needles and give them a sensible motion.

Fourth—the want of a sudden source of magnetic power, for Italics
the bringing of a magnet into position either from a distance or
by inversion, or the making of an electromagnet, both require two words
too much time to allow of the small differences I want being
rendered sensible by them.

But in respect of these points, consider whether experiments
on the following principles may be at all likely to be useful. If,
considering the reasons before given there be the least hopes of (13944)
finding the time, these hopes ought to be verified or exhausted.
Can that be done thus? Italics in original

The objection being fourth above would be strongest if we the
attempted to measure the time occupied by the transmission of
magnetic force across the space between A and B, *bringing a magnet*
up to A from a distance for the time required would be considerable,
and unless the time of propagation were almost as much, the ;
latter could be hardly distinguished. If instead of such a process, hardly be
we had a bar magnet and inverted it—the time of inversion Italics at A
might be made much less than the travelling time in the former
case. Still, it is not likely to be less than the time of propagation, Italics
and if it were not, the latter could hardly be distinguishable
in the first instance. The time required to make or invert Italics
[electrically] an electromagnet might probably be rendered much
less than either of the former periods—still it is very sensible,
and if that of propagation is not greater, the latter would hardly
appear, especially as the electromagnetization is not a sharp but
a progressive operation.

The lateral force of an electric current is of the same nature
as the magnetic force, and is as I think the same force, and its
generation is as rapid as the production of the electric current
itself. Now by the combination of the electric spark and the
Voltaic battery, we can make that as sudden as light almost, and

FARADAY'S FIELD THEORY 111

if we had only means of appreciating the lateral or magnetic
action at a distance, we might I think count upon the generation
of the force as instantaneous.

Now the distant appreciation belongs to the third objection (13947)
—how can we meet this? If the force were strong enough to
generate an electric current in a wire at the place B, that
generation would be as instantaneous as the generation of the
force at A, and then perhaps the time of the propagation from Italics
A to B, whether through space, air, or water, or other media,
might appear, being separated. But a new difficulty is, how are
we to perceive or know without loss of time the generation of the
current at B? Common Galvanometers would be far too sluggish l.c.
to serve as indicators in such a case.
Cannot a galvanometer be made of a magnet so small and fine
as to have only a grain, or the tenth or hundredth of a grain,
in weight? The magnet will be strong—even if two are used to
make an astatic arrangement [one indifferent to the external astatic
magnetic field of the earth], I think that the weight of the whole
need not more than 1/100 of a grain. Would ~~not such a~~ magnet electric
shake instantly 'B' a very small current found it? And it is the by A
first shake we desire to know of. The suspension must be fine
cocoon silk and long. The wire round it may be either a single round
coil or helix. It is worthwhile making such a galvanometer two words
needle and trying it by ordinary feeble currents. (15405)

Now as to arrangements that will include distance: —
if A, B, C, be three rings of wire (copper), and an
electric current set up in one will induce currents [in]
in the two others. If a current were sent through A,
would not its commencement induce currents in B
and C, and if time were concerned would not the Italics
current in B occur first, i.e. be separated by an current
instant from that in C?

The current in C would probably be strong enough because of
its central position, though C is more distant from A than B is: —
besides, if it were not strong enough, C might be a helix of
several coils. (...) end of paragraph

Later Faraday thought of using reflections of light off the two galvanometer needles in the circuits B and C to compare them. He made many other improvements, and tried the experiment and variations of it. None of the experiments gave positive results. The finite time of propagation of the electromagnetic field was not

Figure 1.

In another passage on page 76 Berkson quotes Faraday as follows:

> Even the difference between common current produced by the discharge of static electricity and voltaic electricity, which Dr Wollaston has pointed out, seem explicable by the circumstances connected with the induction of electricity from these two sources.

Faraday wrote in paragraph 76 of his *Experimental Researches in Electricity:*

> Even the differences between common and voltaic electricity, when applied to effect chemical decomposition, which Dr Wollaston has pointed out, seem explicable by the circumstances connected with the induction of electricity from these two sources.

That these two statements are different is obvious and the differences are of some importance. To attribute the phrase 'common current' to Faraday at precisely the time he was beginning to doubt seriously that there was such a thing as a current is to do mortal injury to the historical record.

To errors in transcription must be added serious errors in description. Here, for example, is Agassi's description of the famous discovery in 1831 of electromagnetic induction by means of a heavy iron ring wound with a primary and a secondary coil (p. 72).

> In August 1831, Faraday tried again the experiments of 1825 and 1828. This time he did not put two wires near each other but rather coiled two long wires on the two sides of a wooden ring. The experiment was thus repeated, again with negative results (*Experimental Researches in Electricity*, section 7). The experiment was repeated again with still longer coiled wires, with an iron instead of a wooden core, and with increased currents. Still no results. 'Many similar experiments' (section 9) led to the same negative result. Still longer wires were employed, the isolation of the wires was better secured. Success at last!

This description serves Agassi's philosophical prejudices perfectly. Here was no empiricist following the dictates of the negative results of his experiments, but a theorist and a metaphysician who had deduced a necessary effect from his metaphysics and theories and who refused to accept the negative results he obtained. So, like Robert the Bruce's spider, he repeated his efforts, changing the conditions, until success was achieved. What better example could one ask for to illustrate the importance of having a coherent metaphysics from which theories and facts could be deduced?

The one unfortunate element in this account is that it is totally at variance with the facts. Faraday did not start with a wooden ring or with coiled wires nor did he fail to obtain positive results on the very first run. It is all nicely laid out in the *Diary* where the entry for 29 August 1831 reads:[1]

---

[1] *Faraday's Diary*, volume 1, p. 367. The only full account I know of the rationale behind this experiment is that offered by Williams [1965], pp. 169 ff. Agassi and Berkson neither refute nor disagree with this account; they simply ignore it although it offers a coherent explanation of what Faraday appears to have been thinking. Berkson does refer to it in passing but never mentions Faraday's work on acoustical figures which Williams argues was an integral and central part of the rationale behind the ring experiment.

> 1. Expts. on the production of Electricity from Magnetism, *etc. etc.*
> 2. Have had an iron ring made (soft iron), iron round and 7/8 inches thick and ring 6 inches in external diameter. Wound many coils of copper wire round one half, . . . [The account goes on to describe the effect on the galvanometer in the secondary circuit when the primary circuit was made and broken.]

Where, then, does Agassi's account come from? In the published paper, Faraday wrote of his success in producing currents by induction, 'These results I purpose describing, ***not as they were obtained***, but in such a manner as to give the most concise view of the whole' (*Experimental Researches in Electricity*, paragraph 45, my italics). The experiments that Agassi describes as leading up to the ring experiment were, in fact, done after it and Faraday uses them in order to illuminate the principles of electromagnetic induction that he has discovered and wishes to describe. I am simply at a loss to offer an explanation of why Agassi presents what is palpably false as a true account unless it be the result of the power of his own preconceived notions of how science in general and Faraday in particular proceeded.

Berkson is guilty of a similar failure to read Faraday closely and attentively. On page 69, he describes Faraday's experiment with what is now called unipolar induction in which Faraday discovered that the lines of force of a revolving cylindrical magnet do not rotate with the magnet. This fact was of crucial importance in the evolution of the field concept for it showed that the magnetic force, whatever it might be, was not tied to the molecules of the magnet as Ampère had insisted. Berkson seems not to appreciate the importance of this effect in Faraday's thinking for later, on page 107, he writes that 'all the effects can be equally well explained by assuming that the lines of force rotate *with* the magnet; the same results follow'. Faraday was not assuming: he knew that the lines of force did *not* rotate and so Berkson's discussion here, like Agassi's, involves historical inaccuracies.

To complete this tragedy of errors, I shall cite passages where the authors do not seem to have read carefully what they, themselves, have written. The most blatant example comes from Agassi. He quotes Faraday: 'It has been allowed, I believe, by all who have experimented on these phenomena, that the similar powers repel and the dissimilar powers attract each other; and that, whether they exist in the poles of the magnets or in the opposite sides of conducting wires.' (p. 56). He then comments: 'No philosopher, before Faraday, spoke clearly of *forces acting on forces*, save perhaps Wollaston (and then not in print).' (Agassi's emphasis). Faraday nowhere uses the word, force, in this passage and anyone who has studied Faraday carefully will know that it is illegitimate to assume that he equated power and force. He was ever aware of the incredible power of words to mislead—this is why he was always so careful when he used the term current for he knew it implied a whole theory of electricity. Later, for example, in trying to describe magnetic lines of force he referred to them as 'floods of power' and it was obvious that power and force were not the same here. On pages 78 and 79 of Berkson, we read 'It could be seen that equivalent amounts of static and voltaic electricity (as measured by the magnet) decomposed the same amount of electricity, thus showing the identity of effect of frictional and voltaic electricity.'

And, 'Finally, Faraday found that the same amount of electricity was required to decompose a given amount of a material, say oxygen, no matter what it is in combination with'. In both cases, we can puzzle out what the author means but it is hardly reassuring to have electricity decomposing electricity or oxygen, rather than a salt, being decomposed.

I have spent an inordinate amount of space pointing out these errors because Agassi has thrown out a challenge to scholarship that I feel cannot be allowed to pass unanswered. Berkson seems to have followed Agassi and before their usages can swell into a trend, it must, I believe, be firmly blocked. In the quotation at the head of this review, Agassi has asked us to take his documentation 'less as a tool of scholarship and more as an attempt to create an air of verisimilitude'. Further, we are to read his work as a historical novel. As an ardent reader of historical novels, I find nothing wrong with this approach except to ask, what is the function of the historical novel in revealing to us historical truth? I believe the answer is simple. The historical novelist is interested in motivation and character and many facets of the past for which we have no documentation whatsoever. The novelist uses the known facts as a background against which the characters move and act. Such works as H. F. M. Prescott's *The Man on a Donkey* are both works of art and beautiful examples of the way in which the novelist can illuminate and flesh out history. But to call a work a novel does not give license to falsify the historical record. I fear that these errors are not the kind that lead to truth. In fact, they create confusion and Bacon would, I am certain, be impatient with them. Nor do I find them interesting or in any way preferable to good old, dull, truths. Surely they are to be condemned, for no appeal to 'entertainment' or 'verisimilitude' can alter the fact that they are misleading and false and serve absolutely no valid purpose whatsoever.

Beyond the question of accuracy, there are some other important matters that deserve discussion. These involve the quotation from Collingwood which, if I read him right, is to be interpreted as meaning that the historian and especially the biographer must attempt, insofar as is possible, to put himself literally in his subject's place. The objection that neither of these works is a biography will not stand, for both authors are attempting to reconstruct the course of Faraday's thought and insofar as they attempt this they are writing biography. Collingwood's remarks are not to be seen as suggesting that the historian ask 'what would I have done in such a case' for the 'I' must be the historical personage, not the historian. This, however, is precisely what Agassi and, to a lesser extent, Berkson do. Agassi sees Faraday as a slum kid with a crippling cockney accent, struggling against bitter odds to 'make it' in the high-toned world of English science at the beginning of the nineteenth century. Here are some typical sentences from Agassi's first few pages:

> Faraday's early years were very hard; he was often hungry and had almost no schooling. He was physically too weak to follow his father's trade (that of blacksmith) (p. 1). The hypersensitive struggling youth, as we shall see, became only more tormented and more struggling as the years went by (p. 19). This is why he encountered so much silent opposition about his views . . . (p. 7). When Einstein said so the public was merely puzzled; when Faraday said so the public was deeply shocked and incredulous [this

> on vibrations without the ether] (p. 5). The numbering [of the paragraphs in his *Diary* and in the *Experimental Researches*] is partly an expression of Faraday's obsessive nature, partly an expression of his sense of mission (p. 9).

And Berkson:

> Faraday's success story is so phenomenal that it is sometimes difficult to think of him as real, especially at our distance from him in time. It is hard to imagine him, in the beginning, as a strong-willed but somewhat pathetic working class boy, trying to learn proper English and to get a job near some scientist (p. 17). Perhaps the most striking feature of Faraday in his later work is that *he felt* a constant burden of struggle for understanding . . . (p. 17, my italics).

Here there is certainly a novelistic quality, but it is of the novels of Horatio Alger with a little touch of Bud Shulberg's Sammy Glick thrown in. Michael Faraday, poor and physically weak, struggles manfully with an incredulous public and shocked colleagues through the years, unsuccessfully, to establish his theory of the field. Always of an obsessive nature, Faraday continued to hurl himself against a hostile world until he was ultimately led to a nervous breakdown about which Agassi writes, 'There is little need, in this post-Freudian age, to interpret Faraday's illness—it is so classic' (p. 293). The final insult came when the Royal Society refused to publish his paper on the relation of gravity to the other forces of nature. He died with only one or two people knowing his real worth and only now can he be properly appreciated. How tragic, romantic and dramatic! One nagging question intrudes: is it true? Here we cannot point to solid evidence and explicitly refute our authors' views but there is some reason to believe that this is a caricature that does a real disservice to Faraday. Let us look at what evidence we do have for Faraday's youth, maturity and scientific 'struggles'. He was poor, for his father was not well and could work only sporadically. Once, he recalled later, he had had to stretch a loaf of bread to last him a week. That this did not affect his physical stamina is shown by his remark to his friend, Benjamin Abbott, that he could strike 1,000 blows in succession with the mallet used for forming the backs of books being bound. This may not be up to blacksmith standards (although I suspect it is) but I doubt if any of us could duplicate this feat. Faraday was not a physical weakling, in his youth or even in his old age. When he was 70, he wrote to Trinity House that he had been able to carry out a trial run at a lighthouse 'by climbing over hedges, walls, and fields' that had been blocked by a heavy snowstorm.

Neither author presents any evidence for Faraday's 'hypersensitivity'. He did resent Lady Davy's snobbery (as did Sir Humphry!) and he was unhappy with the role of valet to Davy that he found himself playing in spite of Davy's assurances that he would not be asked to perform these duties. He resented being accused of plagiarism and he constantly tried to set the record straight on priorities of discovery, sometimes finding that an effect that he thought he had discovered had, in fact, been reported by someone else previously to his own publication of it. This seems to me both normal and generous and not bordering at all on the neurotic as implied particularly by Agassi. Why the numbering of the paragraphs in his researches and his *Diary* should be called obsessive completely escapes me. Faraday had a tendency to work on many things at once. He always had a poor memory. By numbering his paragraphs he created for himself a very handy reference system that allowed him to keep both his thoughts and his results in order. The system worked marvelously well. It allowed Faraday to move easily back and forth between the various series of his *Experimental Researches* with a minimum of distracting references. In the *Diary*, it permitted him, likewise, to collect new evidence for old ideas under appropriate paragraph numbers so that any problem upon which he had ever worked could be re-examined in the light of the latest research on the subject. Faraday's practice, in fact, is a model of the way in which an experimentalist ought to keep his laboratory notebooks and report his findings since it does permit such ease of revision and reference.

Was Faraday's life a constant struggle for recognition that he felt was denied him? Again, there are problems of evidence. If the opposition to his views was silent, how do we know that it was there at all? How does Agassi know that the public was shocked and incredulous when he *ad libbed* his thoughts on ray vibrations? And can Berkson offer any evidence that Faraday did feel a constant burden of struggle for understanding? I know of none that would support either author in this contention. Faraday was content to work alone; he positively shunned students because he felt he could not work with students around. Agassi again mis-states the situation when he says that Faraday 'viewed his being childless as the greatest misfortune next to his having no pupils' (p. xii). Of course, he wanted his views to be understood and perhaps even accepted but nowhere is there any sense in Faraday that he was bitterly disappointed when they were not. Agassi quotes a paragraph that he fails to appreciate and which is our only source for Faraday's feeling towards his own ideas. It is worth citing here. It is the last paragraph in his introduction to the collected papers that make up the first volume of his *Experimental Researches in Electricity*.

> Other parts of these researches have received the honor of critical attention from various philosophers, to all of whom I am obliged, and some of whose corrections I have acknowledged in the foot notes. There are, no doubt, occasions on which I have not felt the force of the remarks, but time and the progress of science will best settle such cases; and although I cannot honestly say that I *wish* to be found in error, yet I do fervently hope that the progress of science in the hands of its many zealous present cultivators will be such, as by giving us new and other developments, and laws more and more general in their applications, will even make me think that what is written and illustrated in these experimental researches, belongs to the by-gone parts of science.

Those who would wish to understand Faraday's character would do well to ponder this paragraph long and well. Is it the remark of a man *struggling* to establish his own theories or *struggling* for recognition of his own originality? Is it, then, a masterpiece of hypocrisy, used by Faraday to mask his feeling of burden? And, if it is neither of these, then how can we account for it?

Neither Agassi nor Berkson pay any serious attention at all to a dimension of Faraday's life that he, time and time again, both said and showed was of the utmost importance to him. Faraday was a Sandemanian—in modern parlance, a 'born-again' Christian and one of the striking characteristics of the born-again Christian is patience with his fellow man. He *knows* that his colleagues are in error and he also knows that the truth cannot be forced upon them. He is content to speak the truth as he knows it and hope that the world will ultimately attend to it. That attitude from his religion, I would maintain, carried over to his science and everything points to Faraday's willingness to let it rest there.

But, then, what of his neurosis that struck exactly at the time he wrote these lines and incapacitated him for the next six years? I cannot agree with Agassi that his symptoms are so classical that they do not even require that one identify specifically what his neurosis was. Psycho-analytical action-at-a-distance, when that distance is across time, is a very dangerous exercise and I am unimpressed by Agassi's analysis or medical credentials. Edward Hare, M.D. of the Maudsley Hospital, London, has recently examined the evidence available on Faraday's mental condition and finds the symptoms suggest a diagnosis of 'temporary ischaemic cerebral attacks in the vertebro-basilar territory'.[1] I hope that Dr Hare's article will soon be printed so that others more competent than I can judge its validity, but I found it a more convincing and scientific analysis than that of 'neurosis'. At least it provides a basis for challenging the interpretation that Faraday suffered a breakdown because of his incessant 'struggles'.

Let us, finally, look at the evidence cited to show that Faraday was ill-used by his fellow scientists. This is the rejection of his paper on gravity. Was this really such an insult? In 1850, the Royal Society had published his twenty-fourth series 'On the possible relation of Gravity to Electricity' in which he had spelled out fairly completely his ideas on the gravitational force and its relationship to other forces, particularly (but not exclusively) electricity. His later experiments (which, incidentally, neither Agassi nor Berkson pay much attention to) described more conditions under which the effects he expected to observe did not appear. Was it really, as Berkson writes, 'pressures from the establishment in the scientific community' that caused him to withdraw his paper or was it, merely, that he recognised the good sense in Stokes's remark that purely negative results had little to offer. We must remember that Faraday was no Semmelweis, forced to stand on street corners in Vienna and pass out copies of his articles to try and get a hearing for his views on childbed fever. He had behind him one of the more prestigious scientific institutions of Great Britain. Had he thought his ideas on gravitation in the later paper really worth publishing, all he had to do was to give a lecture on the subject and publish the results in the newly founded *Proceedings of the Royal Institution*. The idea of Faraday as a rebel confronted by the orthodox, conservative 'establishment' just won't wash; Faraday, himself, was a member in good standing of this very establishment!

If my version of Faraday's character and position in the nineteenth-century scientific community is correct, then we are faced with a serious problem. Why was it that his ideas were not recognised, his theories not taken seriously and his vision of field theory not developed until after his death. One part of the answer to these questions seems to me very clear. It is that Faraday's writings are extremely difficult to understand. Until the 1840's he successfully hid his metaphysical assumptions from his audience and tried to present both his results and his theories in language that would not require metaphysical explanations. Failure to understand Faraday was not due to hostility or even commitment to rival theories; it was, in the main, due to Faraday's own obscurity of exposition. In 1840, Faraday himself acknowledged this fact.

---

[1] Edward Hare, M.D., 'Michael Faraday's Loss of Memory'. Dr Hare kindly sent me a mimeographed copy of this article which, I am certain, he would also be glad to make available to anyone desiring to consult it.

> A man who advances what he thinks to be new truths, and to develop principles which profess to be more consistent with the laws of nature than those already in the field, is liable to be charged, first with self-contradiction; then with the contradiction of facts; or he may be obscure in his expression, and so justly subject to certain queries; or he may be found in non-agreement with the opinions of others. The first and second points are very important, and every one subject to such charges must be anxious to be made aware of, and also to set himself free from or acknowledge them; the third is also a fault to be removed if possible; the fourth is a matter of but small consequence in comparison with the other three; for as every man who has the courage, not to say rashness, of forming an opinion of his own, thinks it better than any from which he differs, so it is only deeper investigation, and most generally future investigators, who can decide which is in the right. (*Experimental Researches*, volume 2, p. 266.)

Note that he adopts the same attitude of patience here as formerly but that he also recognised that his language and descriptions of electrical and magnetic processes were not as clear as they might be. How obscure were they? I think it fair to say that no one before Williams published his biography in 1965 had made coherent sense of his theories and showed how his ideas and experiments hung together in one, theoretical, whole. Williams was able to do this only because he had been able to use the whole of Faraday's laboratory *Diary* and masses of letters and other manuscripts that provided hints and clues with which the course and content of Faraday's thought could be reconstructed. How could it be possible for Maxwell or Tyndall or Helmholtz or Sylvanus Thompson or other commentators to see what Faraday, himself, had only dimly glimpsed and had expressed even more obscurely? Where Agassi and Berkson would see a conspiracy of silence or outright hostility, I must insist there was imprecision, haziness and concepts only half thought through and expressed. In short, Faraday is a *very* difficult scientist to read and make sense of.

This brings me to the last and most serious of my criticisms of Agassi and Berkson. They do make sense of Faraday but they make a hash out of his scientific endeavours. The reason for this is elementary. They both, quite rightly, single out Faraday's idea of the field, of the ability of space itself to be the vehicle of force, as his most daring and important conception. It is this idea that was to revolutionise physics and it certainly ought to occupy the centre of the stage. But both authors are far too impatient. They would have Faraday holding the idea that space, alone, can transmit force from the very beginnings of his researches. Agassi would date Faraday's realisation that 'space is filled with forces that are not bound to particles' to 1821 (if I read his somewhat ambiguous sentence on p. 59 correctly). Certainly, Agassi would insist that the field theory (*i.e.* strains in empty space) was characteristic of Faraday's thought from 1831 and the induction ring experiment on. Berkson also would appear to accept 1831 as the crucial date for the birth of Faraday's field theory. He remarks immediately after recounting the course of Faraday's experiments with electromagnetic induction, 'Faraday first thought of applying his field theory to electrochemistry' (p. 74). By doing this, both authors neglect great masses of evidence to the contrary. Specifically, both authors are at pains to show that Faraday's acceptance of Boscovich's theory of point atoms was both early and temporary and that he abandoned this theory as soon as he had possession of the true field theory. This leads them into real difficulties. Each author spends a good deal of time and space examining Boscovich to show that his ideas were at variance with Faraday's. But Faraday could not read Latin and did not read Boscovich. Nor did he have to. All that he got from Boscovich was the idea of point-atoms, surrounded by forces. How these forces existed did not trouble Faraday in his early years. What seems ever to have been before his eyes was these particles and their associated forces. Faraday was a chemist and he had always to keep in mind that chemistry deals with singularities that were called chemical elements. Combinations of Boscovichean point-atoms provided those singularities and he was not prepared to abandon them. During the 1830s, when he was working on electrochemistry and electrostatics, he dealt constantly with these particles and their ability to undergo strains. Because Agassi and Berkson wish to leap immediately to his later views of the field, they are either at a loss to follow his ideas or they invert them. Thus, Faraday did not 'apply his field theory to electrochemistry'. Rather, it was his discovery that, in electrochemistry, there did not seem to be action-at-a-distance, that led him to his electrostatic lines of force. (See Williams [1965], pp. 191 ff. for a full discussion with ample documentation of this point.) Agassi is led into even greater absurdities. Because he has discarded Boscovich, he makes a complete mess of the 'Thoughts on Ray Vibrations'. In the first place, he confuses the circumstances of this off-the-cuff series of remarks in 1846 with the earlier 'Speculation' of 1844 on electric conduction and the nature of matter in which he first, publicly, insisted upon the advantage of Boscovich's atoms over the conventional billiard balls. The 'Thoughts on Ray Vibrations' not the 'Speculation' as Agassi would have it were *ex tempore* remarks delivered after Charles Wheatstone had fled from the Royal Institution and left Faraday to deliver his Discourse to an expectant audience. The atoms or particles Faraday discusses are clearly Boscovichean, although Agassi denies this. 'Unlike Boscovich's atom and like Kant's', Agassi writes, 'Faraday's is of infinite size'. But Boscovich's atoms clearly are of infinite size as reference to his famous curve will show for there the attractive force of gravity specifically extends to infinity. And it is this force that Faraday suggests may answer as the carrier of the vibrations of light. 'The lines of weight or gravitating force are, certainly, extensive enough to answer in this respect any demand made upon them by radiant phenomena; . . .' (*Experimental Researches*, volume 3, p. 450). Why do Agassi and Berkson not quote this passage? (Berkson explicitly excludes it!) I suspect because it would conflict with their conclusion that Faraday had established the identity of electric current and light. This identity Agassi now christens 'his electromagnetic theory of light' for which there is absolutely no evidence. Similarly, in his discussion of specific inductive capacity, Berkson can write 'when many particles are strung along a line of force' without realising that the strained particles *are* the line of force. The whole series of great discoveries and theoretical insights that make up the first volume of the *Experimental Researches* simply make no sense within the framework of 'lines of force . . . considered as an independent physical existence' (Berkson, p. 87). (See Williams [1965], chapters 6 and 7.)

Because there is no evidence that Faraday held this view of the lines of force, both authors are forced into rather contorted postures. Thus Agassi (p. 250), 'he even suppressed his idea that there was no important difference between the material medium and the nonmaterial medium' (evidence?). 'He discusses field in nonempty space—refusing as yet to discuss fields *in vacuo*' (evidence for the refusal?). The reason there is no mention of the field *in vacuo*, I submit, is because Faraday had not yet had occasion to reflect on the problem. Up to the time of his discovery of the rotation of the plane of polarised light by a magnetic line of force, he had no reason to doubt that lines of force were lines of strain in the force-particles that he had used with such success since at least 1819. It was the realisation that the 'polarity' of the 'strain' he thought he had detected with the 'magnetisation of light' lay in the line of force, not the particles of the glass through which these lines passed, that began his thinking on the conduction of the magnetic line of force through empty space. It was this idea which occupied him during his last ten years of research and which he finally clarified sufficiently to permit him to see what must occur. But what he left behind was a dual system. Electricity was polar and associated with the particles of matter with which electricity seemed necessarily connected. Electrostatic lines of force had ends; the line was the strained train of point-particles. Magnetic lines of force were continuous and did not pass from particle to particle. It was the understanding of this fact that led Faraday to his greatest leap of abstract thought, that space itself carried the magnetic strain. This came at the end of his life and was the culmination of thirty years of sustained thought.

By ignoring and denying this development of Faraday's ideas, both authors again are false to history and to art. If Faraday began his career with his ideas already formed, there is little drama in his life. What makes it both exciting and inspiring is the evolution of these ideas and the constant interplay between his fantastically rich imagination and his experimental testing of his ideas.

It is time, now, to comment on the title of this essay review. I mean the question quite seriously and I shall now answer it with a resounding 'NO!' Let me generalise at my own peril. Philosophers tend to be interested in ideas, their logical connections and their logical consequences. They do not seem to find it very interesting to ask where ideas came from, how they developed and how they were interpreted by others who claim to have been influenced by them. They are, therefore, at their best when analysing a system; as we have seen, they are at their worst when trying to account for the evolution of one. I suspect that they cannot really follow Collingwood's advice and try to think as their subject does. Again, they tend to answer the question, what would I have done in so-and-so's place?—which is quite a different thing.

There is an even more specific philosophical point at issue. Both Agassi and

Berkson seem to have adopted Sir Karl Popper's doctrine of refutability by suggesting that one can write history as one does science by making (refutable) conjectures or statements that have some probability of being found true or, at least, of withstanding criticism. We have already quoted Agassi to the effect that he wishes his book will provoke 'those thoughts that may show me in error and thus enlighten me'. This is all very well for science where conjectures can stimulate new experimental programmes that reveal new facts but it will not do in history. History is an inductive science. There are a finite number of facts that pertain to Faraday available to the historian. I feel secure in saying that it is highly unlikely that we will discover a new manuscript or letter that will radically change our knowledge of what Faraday did and wrote. That being the case, any attempt to understand Faraday and discover him, in Collingwood's sense, by getting inside his mind and following his ideas *must* take the whole body of historical facts on Faraday into consideration. The historian is not free to pick and choose those facts that suit his thesis, no matter how original or fascinating that thesis may be. Nor is it legitimate to make statements for which there is no evidence whatsoever just because one wishes to establish an idea that is essential to one's analysis. If one does, the result is not falsifiability and progress, but falsehood and bad scholarship.

L. PEARCE WILLIAMS
Cornell University

PHILOSOPHISING ABOUT GALILEO*

1 Shapere's *Galileo* begins with a chapter entitled 'Galileo and the Interpretation of Science', which consists of two parts: a statement of the aim and method of the book, and a sketch of Galileo's life. The stated aim is to study some of Galileo's ideas and certain aspects of his method in order to understand the rise of modern science, Galileo's role in it, and thus the nature of scientific development in general, and ultimately the nature of the scientific enterprise itself. The method he will follow, Shapere tells us, is to examine some of Galileo's statements and arguments in the light of a number of interpretations, which will thereby be critically examined. The major interpretations he describes are: the one propounded by Ernst Mach, and to which physics textbooks typically pay lip service, claims that Galileo was an empiricist and a revolutionary innovator; the one developed by Pierre Duhem claims that Galileo was the continuator and culminator of a medieval tradition of Aristotelian critics and commentators; the one elaborated by Alexandre Koyré claims that Galileo was an aprioristic intellectualist (a 'rationalist' is Shapere's word) and a Platonist.

In the second chapter, entitled 'The Intellectual Background', the author lays the groundwork for his examination of Galileo's originality, which is a feature dealt with in each of the three prevailing interpretations. Shapere does this by summarising what he thinks are the relevant views of Plato, Aristotle, the fourteenth-century impetus theorists, and previous astronomical thinkers.

The third chapter is entitled 'The Early Development of Galileo's Thought' and includes a section on his Pisan period and one on his Paduan period. The 'Pisan Period' is a clear exposition of the main ideas in Galileo's *On Motion* and a plausible interpretation of it as being partly Aristotelian, partly Archimedean, and partly impetus-theoretic in intellectual content. The 'Paduan Period' notes that Galileo's ideas at that time corresponded basically to those found in *Two New Sciences*, which raises the question of how he arrived at these from his Pisan theory; Shapere then concentrates on the problem of how Galileo originally arrived at the law of squares, discusses various available interpretations, and concludes that the historical evidence is inconclusive.

Chapter four, entitled 'Galileo and the Principle of Inertia', discusses the evidence from the *Dialogue*, that from the *Discourse*, and some problems about determining to what extent Galileo approached the modern conception of inertia. As regards the *Dialogue*, Shapere argues that Galileo tries to establish the following five propositions: (1) 'anything but rest or circular motion is contrary to the orderliness of the universe' (p. 87); (2) 'natural circular motion is uniform and perpetual' (p. 89); (3) 'the only purpose that might possibly be attributed to straight-line motion is that it could be used to create or restore orderliness' (p. 91); (4) 'if non-circular order-restoring motion does indeed exist, it would be accelerated and rectilinear' (p. 93); (5) 'really, what appears to be natural accelerated straight-line motion is uniform and circular after all' (p. 98). Concerning the *Discourse*, Shapere has the following argument: although Galileo's analysis of the parabolic trajectory of projectiles uses the idea of a rectilinear horizontal natural motion, in his reflections on the problem he claims that the rectilinearity of the horizontal component is only an approximation, since he says that such a component follows the curvature of the earth; hence, argues Shapere, Galileo's inertial component is really circular, from which he concludes, interestingly, that Galileo's commitment to Copernicanism is primary even in the *Discourse*. Shapere's final conclusion is that, although Galileo's conception of natural motion has something in common with 'inertia', it is far from approaching the latter but is rather the culmination of traditional ideas (p. 87), and that the Drake–Koyré dispute about whether Galileo understood the essence of 'inertia' is semantical (p. 125).

---

* Review of Dudley Shapere [1974]: *Galileo: A Philosophical Study*. Chicago: University of Chicago Press. £4·90. hardback £1·50 paperback. Pp. ix+161.

The last chapter, 'Reason and Experience in Galileo's Thought', is basically a criticism of Koyré's Platonic, 'rationalistic' (*i.e.* aprioristic) view of science as it is based on Galileo's work. Shapere argues (rightly) that his use of the Socratic method does not imply epistemological innatism or apriorism since Galileo claims to be reminding us of experience we may have had already but forgotten. As regards Galileo's mathematical approach, Shapere argues that he was somewhat ambiguous about its power to give a full understanding of nature; for Shapere it is more important that such a mathematical approach is no less anti-Aristotelian than anti-Platonic, since both Plato and Aristotle agreed that mathematics cannot give an adequate account of the physical world, and disagreed only on their reasons for this view. He concludes that Galileo's practice was not and could not have been aprioristic ('rationalistic'); that it was partly original and partly traditional (medieval); that it was mathematical as well as empirical, not in the simplistic Machian sense, but in an hypothetico deductive manner.

2 Galileo scholars can and will find fault with this book for reasons such as the following. In his discussion of Galileo and inertia, Shapere claims that one proposition that the *Dialogue* tries to establish is that 'if non-circular order restoring motion does indeed exist, it would be accelerated and rectilinear' (p. 93). The following passage is then presented as containing one of Galileo's supporting arguments:

> Every body constituted in a state of rest but naturally capable of motion will move when set at liberty only if it has a natural tendency toward some particular place; for if it were indifferent to all places it would remain at rest, having no more cause to move one way rather than another. Having such a tendency, it naturally follows that in its motion it will be continually accelerating. Beginning with the slowest motion, it will never acquire any degree of speed without first having passed through all the gradations of lesser speed—or should I say of greater slowness? For, leaving a state of rest, which is the infinite degree of slowness, there is no way whatever for it to enter into a definite degree of speed before having entered into a lesser, and another still less before that. It seems much more reasonable for it to pass first through those degrees nearest to that from which it set out, and from this to those farther on (pp. 93, 94).[1]

In analysing this passage Shapere's main point is that 'the argument does *not* prove that a body moving in natural straight-line motion will continually accelerate; all it shows (assuming the Principle of Continuity, of course) is that, *if* the body is to arrive at any particular finite non-zero velocity, it will have to pass successively through all lesser velocities; it does not show that when any given speed is reached, the body will continue to increase (rather than maintain or decrease) its speed beyond that until it arrives at the goal' (p. 94). However, Galileo is in fact inferring *only* what the argument allegedly shows: he is *not* drawing the consequence that allegedly does not follow. This can be seen from Galileo's original Italian text, where the conclusion reads that the body '*nel suo moto si anderà continuamente accelerando*'.[1] Shapere apparently was misled by Stillman Drake's translation of '*continuamente accelerando*' as 'continually accelerating' and interpreted 'continually' to have a meaning distinct from 'non-discontinuous', which it sometimes has. In other words, Shapere is distinguishing two kinds of accelerations: non-discontinuous and non-stopping. He thinks that Galileo is in this passage talking about both, whereas Galileo is only talking about the former.

---

[1] Quoted by Shapere from Galilei [1953], p. 20.

However, such knit-picking would be basically misconceived. For though it is part of Shapere's aim to be as scholarly as possible (p. x), and though scholarly accuracy is a general methodological duty, nevertheless scholarship is not the animating force behind Shapere's work. Moreover, in attempting a synthesis of recent Galileo scholarship, the book does something valuable which too many specialist-scholars too often fail to do, preferring instead the comforting safety of exploring small and fine points. Perhaps scholars who have not produced such syntheses lack the right to make any serious objection to those who have; and to say this is not cheap *ad hominem* reasoning, but rather an attempt to make a distinction between atomistic, analytical scholarship and wholistic, synthetic scholarship, and to suggest that different criteria of evaluation apply.

3 It is at this point that general historians (of science) may enter the scene; for they can and will object to the book for reasons like the following. In his discussion of Galileo's intellectual development, Shapere points out, following Duhem, that there were only three steps left to arrive at an inertial physics from Buridan's impetus theory: 'first, the conception of the possibility of motion in a void or vacuum; second, the abandonment of the idea that there is circular impetus as well as rectilinear; and finally, the abandonment of the idea of impetus as an internal force distinct from the body and maintaining it at constant speed' (pp. 50–1). Hence, he concludes, 'if this is a true historical picture, the Impetus Theory would thus stand as a transitional phase between these two traditions' (p. 51). Now, at this point a historian would try to determine whether, or to what extent, 'this is a true historical picture'. Instead, Shapere proceeds to ask, 'If the Impetus Theorists were in this respect so close to modern physics . . . why did the transition—if it did, as a matter of historical fact, take place *via* the three steps mentioned—take so long?' (p. 52). The parenthetical expression in this quotation may be an indication of the attitude that questions of historical fact are themselves 'parenthetical'; it may also be a natural rhetorical device to express ambiguity as to which of two distinct questions are being answered: (1) Why did the transition along these three steps take so long? (2) Why did *not* the transition take place along these three steps? What Shapere

[1] Galilei [1890–1909], p. 44.

proceeds to do is to discuss the intellectual obstacles to the transition in question (pp. 52–6). The expressed ambiguity hides the fact that the question that has really been answered is: Why did not the transition take place along these three steps, or *if* it did, why did it take so long? In other words, the intellectual obstacles presented by Shapere are indeed both plausible and likely historical causes, but they are either dilatory or preventive causes, and he gives virtually no evidence or even an argument to exclude the preventive cause alternative. It is not, of course, *my* task here to show that the transition did not take place, since my point is that Shapere has not shown that it did take place; worse, he has not even given any evidence for it; he has only given the above-mentioned ambiguous argument, and hence not really argued, but rather insinuated that modern physics originated from the impetus theorists.

A second example is the anti-Koyré argument where Shapere tries to show that Galileo's practice was not 'rational' because it *could not* have been (pp. 132–3). I do not intend to give the impression that this is Shapere's only argument in this context, for he also gives evidence (rightly, I believe) that Galileo's explicit attitude toward Platonism was ambiguous and evasive and that his actual practice was not in fact Platonistic-rationalist (pp. 131–2). However, the apparently 'philosophical' argument mentioned above will offend historians' sensibilities, though it is doubtful whether they will be able to give a reasoned criticism of it and show that it is self-inconsistent or circular, or both. In fact, Shapere's premise that Galileo's practice could not have been rationalistic is to be justified either *a priori* ('rationalistically') or *a posteriori* (empirically). He hints at an aprioristic justification when he italicises the 'could' and thus indicates that he is referring to a logical or conceptual possibility; and he hints at an empirical justification in saying that 'the history of science and philosophy has not dealt kindly with' (p. 144) rationalism in the sense under discussion. At any rate, the aprioristic justification would itself be rationalistic and thus be doing or trying to do what Shapere's conclusions claims to be impossible; whereas the empirical justification would have to be based on historical case studies showing, for example, that Galileo's practice was not in fact rationalistic, which is precisely what Shapere's present argument tries to show.

Once again, however, such fault finding would be basically misconceived. Shapere's book may be about historical events, but it is not a historical work; its animating principle is not the historical sensibility, and thus any violence to it detracts little from the book's intrinsic value.

4 Philosophy ought not to be equated with either scholarship or history, however much one may love both. A philosophical work ought to be judged by philosophical standards, and Shapere's book is 'a philosophical study'. Now, some philosophers may ask themselves what is the relevance of such a study. The answer is that philosophy at its best is not theoretical and abstract but applied and concrete. The present work is one in the philosophy of science. The philosophy of science may be conceived either as the philosophical study of contemporary scientific knowledge or as the philosophical study of past scientific achievements. The former centres around contemporary science-textbooks, the latter around past scientists. There is certainly no good reason for regarding the former approach as superior to the latter: the latter need not be a duplicate

history of science, any more than the former need be popular science; nor need the second approach degenerate into psychology, any more than the first need degenerate into linguistics, library science, communication theory, or pedagogics. What seems undeniable is that philosophical reflection is an activity that can be exercised irrespective of the objects of study. In a sense, the two approaches are distinct in a complementary way. For contemporary textbooks are paradigm examples of scientific knowledge, whereas past great scientists are paradigm examples of scientific rationality. So what we have is the study of scientific knowledge and the study of scientific rationality. Now, scientific knowledge and scientific rationality are *prima facie* two more or less equally important elements of science; hence a comprehensive philosophy of science cannot limit itself to either one. Nevertheless, one cannot talk or think about everything at once, and so, for some purposes, one may limit himself to the study of scientific rationality. Shapere's book is then a work in the theory of scientific rationality. As such it is to be welcomed since it helps to correct the prevailing over-emphasis of work in the theory of scientific knowledge.

In studying scientific rationality as we find it in Galileo, Shapere realises that it is likely to contain a number of distinct elements, rather than being a monolithic unit. So he selects three distinct aspects of Galileo's work: the laws of falling bodies, his ideas on 'inertial' motion, and his 'scientific method'. These three elements are well chosen since they are traditionally and still widely regarded as Galileo's great contributions to science. If we accepted these three contributions as great, epoch-making, and paradigmatic scientific achievement, then we could study Galileo at work in these three areas, try to determine whether or how he is aprioristic, or empiricist, or mathematicist, or Platonist, and thereby learn what makes a good scientist and understand the nature of scientific rationality. Shapere follows this approach for the case of the laws of falling bodies (pp. 78–86). However, he reaches no conclusion here about the nature of scientific attitude because he thinks that the evidence is insufficient to enable one to determine how Galileo originally arrived at the law that the distance fallen varies as the square of the time.

Unfortunately, Shapere's study of Galileo's ideas on inertia seems no more conclusive. In fact, though he is willing to say rather clearly and explicitly that and how Galileo arrived at his principle of natural circular motion, Shapere also points out that such a principle not only never became a part of modern science but misses the main point of modern physics. According to Shapere, Galileo tried to deduce natural circular motion from the conception of the orderliness and perfection of the universe in an attempt to seek a deeper and more consistent cosmology than Aristotle's. Presumably then, in this case, we know the method used to arrive at a scientifically worthless idea. But this seems to inform us about scientific irrationality rather than rationality. This is especially true since Shapere argues that virtually all of Galileo's justifications of his circularity ideas are logically invalid, so that in this aspect of his work Galileo would be a paradigm example of irrationality, in a double sense, namely that his conclusion as well as his arguments are wrong.[1]

---

[1] This problem cannot be dismissed by saying that all that Shapere is trying to do is 'to find out the facts' about some of Galileo's thinking, and that philosophers can thereby reach any conclusion they want. For to say this would be to regard the book as a scholarly or historical work, which interpretation would be neither accurate nor generous.

The problem with Shapere's examination of the third important aspect of Galileo's work is of a still different nature. It is to be found in the chapter entitled 'Reason and Experience in Galileo's Thought', which is a discussion of the latter's explicit ideas and reflections on scientific method, as distinct from Galileo's method *per se*. In other words, it is an examination of Galileo's so-called philosophy of science, as distinct from the method inherent or implicit in his science. Unfortunately when Shapere tries to extract a philosophy of science from Galileo, it evaporates: allegedly he does not have one, if we mean a systematically and consistently worked out theory of scientific rationality. For Shapere argues that Platonistic rationalism is not an element of his philosophy (pp. 132–3), and 'concerning the limits (if any) of mathematical treatment of nature, it appears that Galileo's position was not clearly formulated, and probably he was not even fully cognisant of the issue' (p. 138). So we cannot put our hands on a Galilean 'philosophy of science'. But even if we could, Shapere recognises and in part argues that its relevance would have to be justified by the extent to which that philosophical theory corresponded to his scientific practice. Hence this part of Shapere's book is as inconclusive as the previous two regarding the nature of scientific rationality. What it does seem to amount to is an argument that it is of little or no value to study Galileo's philosophy of science. And someone who seriously proposed such an argument would have to face the (possibly rhetorical) question of whether there does not exist a better argument for the same conclusion, namely silence.

5 I do not think, however, that the situation is as desperate as it has appeared so far. In fact, suppose we agree with Shapere that there is no way of establishing from the available evidence how Galileo first arrived at the law of squares. It turns out that this lacuna is largely inconsequential and that the evidence is not lacking for what is really needed. What we need is an explanation of Galileo's sources concerning the law of squares. In fact, Shapere is right in saying that 'the importance of the controversy between these three alternative interpretations of the sources of Galileo's correct law of falling bodies must be recognised: for each of the three possibilities lends credence to a different interpretation of the rise of modern science' (p. 86). But if this argument is to be valid, the source of Galileo's law of falling bodies cannot be his method of first arrival at the law. For the fact that a particular scientist (however important) arrived at a particular law (however important) in a certain way would give little support to the generalisation that modern science as a whole originated that way. On the other hand, this generalisation would be *significantly* supported if we could assert that a particular important scientist 'arrived at' a particular important law *because* he proceeded in a certain way. 'Arrived at' here would mean 'was able to arrive at' or 'was successful in arriving at' or 'worked successfully with'. Now, the kind of evidence needed to explain Galileo's success with the law is evidence relating to what he did with it *after* he became convinced that it was true, of which there is an abundance. Shapere himself refers to some of it and describes it as 'searching for a proof of the relationship' (p. 81). This description is correct at a phenomenological level, but the search for a proof can be given a deeper interpretation by showing it to constitute, as I have done elsewhere,[1] a search for an explanation.

---

[1] Finocchiaro [1973].

Turning to 'inertia', can the methodological inconclusiveness of Shapere's account be remedied? I believe it can if we modify his approach slightly. His approach seems to have the following elements: (1) concentration on the context of justification, as distinct from the context of discovery; (2) comparison of the historical material with present-day textbooks; and (3) logical analysis, that is, reconstruction and evaluation of arguments.

Now, there is nothing wrong with emphasising the context of justification as long as one is aware that it is *conceptually* distinct from the context of discovery and as long as the conclusions one reaches are not inappropriate but rather pertain to the context one has examined. Moreover, the conceptual distinction between the two contexts does not mean that they are always separate in reality and in practice; in fact, Galileo's justification of natural circular motion in the *Dialogue* is a beautiful example of such coincidence: his beliefs did in fact originate from the reasons contained in arguments like the ones examined by Shapere.

Second, to compare Galileo's ideas about natural motion with later ideas or present-day knowledge is unobjectionable if one makes the comparison in order to have some contrasts to understand better the content of those ideas. What would be problematic is to conclude from this contrast that these ideas were unscientific or that Galileo was in this instance a bad scientist; but Shapere draws no such conclusions. Instead, I believe he would agree that such conclusions are to be supported by an analysis of Galileo's own evidence and the justifications he put forth; this would explain, in part, why Shapere tries to reconstruct and evaluate Galileo's arguments—the third element of Shapere's approach.

Logical analysis would seem to be ideally suited in a case like the present one. For here the ideas did not (at least allegedly) become part of modern science and thus one cannot ground one's philosophical conclusions on a method-explanation of success, as in the case (however problematical) of Galileo's law of squares; moreover, the origin of ideas here happens to coincide with their justification. Therefore, those who are dissatisfied with the methodological inconclusiveness of Shapere's account here, cannot easily blame it on the method of logical analysis; instead they can try to modify it.

One modification concerns the application of the concept of logical validity in the sense of inconceivability of premises being true and conclusion false. If Galileo's arguments are to be evaluated in terms of such deductive validity, then greater care should be taken in their reconstruction. For example, they should be reconstructed from the original text, not from a translation. The fairness of this requirement is supported by the problem mentioned in another context earlier, pertaining to Galileo's argument for the continuity of acceleration in rectilinear order-restoring motion. But this requirement ought to sound intrinsically plausible to philosophers nowadays since it is just another way of stating the interconnection between logic and philosophy of language. My point is not that deductive validity is a language-dependent concept, but rather that since it is such a rigorous criterion of evaluation, the deductive validity of an *actual* argument is likely to be dependent on how exactly it is stated.

Another modification would be to require that the conclusion of a given argument be identified with as much preciseness as the concept of deductive validity demands for the connection between the premises and the conclusion of an argument. So it is not at all a peripheral issue what exactly is the conclusion of Galileo's argument about the actual path of a falling body: whether it is Shapere's Proposition 5, namely that 'really, what appears to be natural accelerated straight-line motion is uniform and circular after all' (p. 98); or whether it is, as I believe the text more strongly indicates, that: *if* the earth rotates, *then* what appears to be natural accelerated straight-line motion is really uniform and circular, *as a first approximation.*

When care is exercised in the reconstruction of an argument, which is a requirement deriving from the rigorous character of deductive validity, it soon emerges that the application of this concept is severely limited. For one discovers that the connection between premises and conclusion stated in the text is one of probability, or plausibility, or mere relevant evidence; in other words, inductive validity becomes the more relevant concept, as long as one understands the latter in a general-intuitive way rather than a specific, technical-formal theoretical elaboration.

A fourth feature of careful logical analysis would be to pay attention to a fuller context. One might say that logical analysis comes into being when, given a proposition, one asks where it came from in the sense of the reasons being offered to justify it. Now, one is doing a half-way job if one stops here and does not ask where the proposition is going, in the sense of to what use it is being put, that is, what further conclusions are being drawn from it. Of the five propositions on 'inertia' attributed by Shapere to Galileo, the first four occur in the context of Galileo's critique of the Aristotelian *a priori* argument for the terrestrial–celestial dichotomy, the argument from the presence or absence of contrary motions. When this is taken into account, the merit of Galileo's arguments would be different since they would be judged relative to the Aristotelian arguments supporting the two (or three) types of natural motion; such relative assessment (which might be an important part of inductive evaluation) would also yield a different evaluation of Galileo's commitment to 'circular inertia'. As for the fifth proposition, careful logical analysis would note that it is justified in the context of Galileo's refutation of the anti-Copernican argument from falling bodies and is explicitly labelled and implicitly treated as a speculative digression.

Thus I conclude that the method of logical analysis, used by Shapere, can be a very useful tool in the theory of scientific rationality, if it is properly carried out. Such a modified use would presumably reveal a different Galileo in his work on 'inertia'.

In regard to Galileo's philosophy of science, I would agree with Shapere that it is important to be cognisant of the theory–practice distinction, and realise that Galileo's philosophy of science and his scientific practice may not correspond. This means that it is inappropriate to reach conclusions about Galileo's actual scientific procedure on the basis of evidence from his reflections about science. Hence, if Galileo does not have a philosophy of science, this creates no problem for the theory of scientific rationality, whose primary concern is the study of scientific practice. All this is correct as a first approximation but needs to be qualified as follows. If Galileo does not have a philosophy of science, then since he is a paradigmatic scientist, that lack must be significant; perhaps the significance is that scientists ought *not* to have a philosophy of science, if for example that lack accounts partly for Galileo's success; or if Galileo succeeded in spite of the lack, then perhaps his lack would not justify a norm. In other words, a complete theory of scientific rationality must say something about the proper relationship between philosophical and scientific activity, for the simple reason that we find a mixture of both in many paradigmatic scientists. I do not have the answer to this problem, but I wish to suggest one possibility that ought not to be overlooked in searching for a solution: it may be that to *found* modern science required philosophical genius, though the continuation of it does not; this would mean that Galileo need not be paradigmatic for scientists in regard to his philosophical sophistication, even if he does have a philosophy of science which was instrumental for his scientific achievement. In this case the possibility would arise, however, that he might be exemplary for philosophers, in that his example suggests that it might be an element of *philosophic* rationality that one has to be a 'scientist', at least in the sense of a student of a specific field. But these matters need not be pursued any further; suffice it to say that the question of the nature and role of Galileo's philosophy of science is not to be dismissed simply because such philosophy does not have a direct connection with scientific rationality, since the connection may be indirect.

I believe that Shapere is right in saying that Galileo does not have a philosophy of science, in the sense of a systematically formulated theory of science. However, since philosophy ought not to be equated with a systematically formulated theory, it would be improper to say without qualification that he did not have a philosophy. This would be as improper as to say (unqualifiedly) that Socrates was not a moral philosopher (in that he did not have a systematically formulated and fully consistent ethical theory). In fact, Galileo no less than Socrates is a philosopher because both are thoughtful doers and practical, action-minded theorists; in both, thought enlightens action, thinking enriches life. Socrates' domain of concern was morality and the good life, Galileo's was natural phenomena, science, and the well-informed life; but such differences are material, not formal. Galileo's way of living an examined life was to pursue science as he did; it would have been no more appropriate (philosophical) for him to appropriate Socrates's own concerns, than for Socrates to have remained in the grip of natural philosophy, or for us to forget our own problems and appropriate those of either one. It is beyond the scope of the present essay to explore and articulate this idea any further; instead I wish to mention another related fact. It is an overwhelming feature of Galileo's works that he has a methodological awareness; that is, he is always engaged, whenever the need arises, in formulating and justifying principles which underlie the particular ideas he expresses and procedures he follows. It is as true that his methodological discussions always arise out of concrete problem situations as it is that they never reach metaphysical depths; indeed Galileo was not a metaphysician. But to lack any significant metaphysical awareness or interest prevents one from being a philosopher only in the mind of those who equate metaphysics and philosophy. Here too, since Socrates would also be excluded, Galileo would be again in good (philosophical) company.

In conclusion it may be said that if we accept Shapere's negative results (about the origin of Galileo's law of squares, about the historical and logical worth of this 'inertia', and about the existence of a Galilean philosophy of science), they need not be a cause for despair, because it is possible to use them for a better and more philosophical appreciation of Galileo. Specifically we see that the theory

Brit. J. Phil. Sci. 26 (1975), 265–273 *Printed in Great Britain*

Reviews

BUKHARIN, N. (*ed.*) [1971]: *Science at the Cross Roads. Papers Presented to the International Congress of the History of Science and Technology*, 1931. *By the Delegates of the U.S.S.R.*; with a new foreword by Dr Joseph Needham and a new introduction by P. G. Werskey. London: Frank Cass & Co. Ltd. £5.25. Pp. xxx+275.

To write that any work of religion (Sutra, Koran, Gospel...) is a book of enormous influence, of seminal importance, of brilliant illumination, is of course to declare oneself among the believers. Non-believers are apt to regard all (or all but one) of such books with indifference at best. The voices which have declared *Science at the Cross Roads* a volume of comparable significance have been those of believers also. As P. G. Werskey points out in his Introduction to this reprint, they were largely the voices of a group of scientific intellectuals of bourgeois origin (J. D. Bernal, J. G. Crowther, Joseph Needham, Lancelot Hogben, for example) whose ample writings have (deservedly) won many readers. *Science at the Cross Roads* had no effect in the Soviet Union; the road taken by most of its authors led them to the prison camp and the execution squad. I have known no Western European write or speak of the book. It has been read (if at all) only by readers of English. Among them, it had no effect upon the generality of scientists or scientific journalists in either Britain or the United States; and though there were in the 1940s (as later) Marxists among historians and historians of science, as well as among scientists, neither general history nor the historiography of science have been profoundly shaped by dialectical materialism and researches of the Hessen sort. Nor, I believe, have British working-class politics. Whether Mr Werskey is also right to claim, after making his point about the book's adherents, and after explaining how these adherents misunderstood the ideological situation in the USSR in 1931 (not to say the scientific situation also), that *Science at the Cross Roads* is 'one of the key documents' in the 're-evaluation' of the 'older view [of some socialists] that technocratic means and socialist ends are not incompatible' will not be examined here. Ideological, that is to say religious, debates (if any) must occur elsewhere. Has the book any intellectual significance?

Of course it is a valuable historical document—as full of ripe quotation as the writings of any Victorian moralist. The context of its creation is neatly put in the essay of Bukharin. On the one hand the old putrescent capitalist economy, its condition of crisis producing 'a most profound crisis in the whole of capitalist culture; a crisis in individual branches of science, a crisis in epistemology, a crisis in world outlook, a crisis in world feeling;' on the other hand,

> young Socialism is arising... its cultural-historical approach dominated by the Marxist outlook... It is not only a new economic system which has been born. A new culture has been born. A new science has been born. A new style of life has been born. This is the greatest antithesis in human history, which both theoretically and practically will be overcome by the forces of the proletariat—the last class aspiring to power, in order in the long run to put an end to all power whatsoever.

Where was science at the cross roads to go, westwards towards the rottenness of Wall Street or eastwards where the dawn was bright? Of course, by implication science becomes different according to its political orientation; bourgeois physics (Einstein) and bourgeois biology (Mendel) would be supplanted in socialist science. As B. Zavadovsky put it (in a paper praised by Joseph Needham):

> [It is] a matter for regret that the modern naturalist, when studying problems of philosophy and the history of natural science, remains unaware that these problems of overcoming, on the one hand, the most reactionary idealistic and vitalistic current of thoughts, and on the other the oversimplified mechanistic positions of vulgar materialism, were not only formulated but solved in their basic and characteristic principles more than seventy years ago, in the classic works of the founders of the philosopy of dialectical materialism, Marx and Engels, and in our own times in the profound works of Lenin.

Or to take a different example, E. Colman writes that 'only planned socialist economy' can present mathematicians with complex practical problems; but (he continues) planning mathematics does not mean stifling creativity—'it means that mathematics will flourish as it can do only with collective labour, the more so as the ranks of the scientists are swelled by new men from the working class, to whom the gates of the universities and research institutes have been opened wide.'

Thus one may argue that *Science at the Cross Roads* has value as a cluster of predictions, justified by a variety of arguments and evidence as well as by appeal to Marxist writ. Such predictions of difference were more commonly applied to Soviet science forty years ago than now, when they have been transferred to the People's Republic of China. Their fate is too obvious to require comment. Socialist genetics in Stalin's hands created disaster; Soviet high-energy physics, Soviet radio astronomy, like Soviet ballistic missiles and Soviet space-ships, do not differ markedly from those of the West. Saluting the achievements of Soviet science, acknowledging the sacrifice that has made it possible, one is nevertheless unable to see, under socialism, a 'new science' distinct from the equally 'new science' of the West; any more than one can see truth in Lenin's prediction that electrification would bring freedom and joy to socialists, misery and slavery to French peasants.

The basis of the predictions contained in this book (see also those of the veteran Soviet physicist A. Joffe) was hope that a rational, benevolent society must succeed better than an irrational, cruel society, hope buttressed by a firm conception of the nature of science as 'refined experience'. Of course a sophisticated socialist thinker did not suppose that every branch of knowledge—say, the theory of numbers—possessed direct practical significance. Rather (as Bukharin wrote): 'It is a question of systems as a whole, of appropriate activity, of chains of scientific truths, representing *in the long run* the theoretical expression of the "struggle with nature" and the social struggle.' But because the socialist experience was of a wholly distinct character from that of capitalism, its science was to be different also. To quote Bukharin again:

> The elemental regularity of interdependencies between economy and ideology, between collective economic practice and the multifarious branches

> of theoretical labour [in a capitalist society] yield place [under socialism] to a considerable degree to the principles of planning . . . [In the U.S.S.R.] all the poverty of the idea that the 'utility' of science means its degradation, the narrowing of its scope etc. becomes crystal clear and apparent . . . we are arriving ***not only at a synthesis of science, but at a social synthesis of science and practice*** (italics in original).

From decrying the idea of 'science for its own sake' or the 'fetishism of [bourgeois] science' it is only a little step to the argument that because no scientific proposition about the natural world can be held to be finally and absolutely true, all such propositions are only ***relatively*** true, that is, true for the societies that entertain them. So Zavadovsky described how 'Darwinism' afforded support for the 'materialist radicalism' of the bourgeoisie against the religious-idealistic ideology of feudalism', concluding 'All these facts prove the socio-historical and class determinateness of scientific theories'. Later, speaking of the need for a 'planned solution of the problems of large-scale Socialistic agriculture and cattle-breeding' he went on, yet more ominously, 'These necessitate the overcoming both of the mechanistic and Lamarckian conceptions . . . and of the autogenetic enthusiasm of the geneticists, who think that the tasks of the Socialist Five Year Plan are covered by the application of the methods of modern genetics and selection, ignoring the rôle and importance of the rest of man's system of social measures . . .'

Whereas the bourgeois scientist conceived of his goal as approximation to truth (while recognising that the process was neither rectilinear nor single-valued) the socialist scientist of 1931 conceived of the scientist's task as that of providing 'solutions' to 'problems' (a view, perhaps now also more commonplace in the West than it was forty years ago). The questions (1) what counts as a problem (2) what counts as a solution were not clearly discussed by the Soviet authors, though they might seem crucial for an analysis of the function of science in society; however, they were sufficiently confident of their analysis to extend it to the whole history of science. 'We have shown', B. Hessen wrote of Newton's *Principia*, 'how its physical content arose out of the tasks of the epoch, which were raised for accomplishment by the class entering into power.' Alike ignorant of Newton's intellectual antecedents, of his character and connections, and of the English society in which he lived, Hessen framed an argument that Carlo Cipolla has, as it were, turned inside out: when Newton lived the bourgeois were coming to power; the bourgeois relied upon ships, guns and machines; all these are things that move and depend upon the laws of motion; therefore Newton created the laws of motion.

'Why did the ancients have no dynamics?' another author has asked (elsewhere). Because they had no firearms, was his answer. Why was Icarus ignorant of aerodynamics? Because he had no aeroplane . . .

A. RUPERT HALL
Imperial College,
University of London

GALILEO [1974]: *Two New Sciences, Including Centres of Gravity and Forces of Percussion.* Translated, with Introduction and Notes by Stillman Drake. Madison and London: University of Wisconsin Press, 1974. Clothbound $14.00, Paperbound $5.00. Pp. xxxix+323.

Professor Stillman Drake, whose diligent work on many aspects of Galileo have stretched over the past several decades, has now published a new English version of what is perhaps Galileo's most significant work. This is actually the fourth English translation of the *Two New Sciences*, but should without question replace the one by Crew and De Salvio, first published in 1914, as the standard English version of this important work. The new translation includes, besides the integral text of the *Discorsi e dimostazioni matematiche intorno a due nuove scienze*, also an appendix containing Galileo's *On Centers of Gravity* and an additional 'day' of dialogue *On the Force of Percussion*, here translated into English for the first time. The volume includes a useful introduction, explanations of various key mathematical and physical terms used by Galileo, rather sparse notes, and a carefully prepared index. In the Preface Drake mentions that this volume is to be followed by two others, one which 'will include documents, fragments, and correspondence relating to Galileo's new science of motion' and another which will contain 'detailed notes to the present translation as well as to the documents and correspondence, and a general biographical and historical appraisal of Galileo's physics'. Thus, we should eventually have from the translator a comprehensive statement of his reconstruction of Galileo's development. The publication of these future volumes will be of value, for they should bring together a broad range of materials not generally accessible to English readers of Galileo's major work. The reader who relies primarily upon English materials has at present nothing comparable to the illuminating notes and commentary to the work prepared in Italian by Carugo assisted by Geymonat, a volume which remains one of the best annotated editions of a major scientific work presently available. It should be noted, however, that the major new interpretative work by Winifred L. Wisan (Wisan [1974]) just published should go a long way towards clarifying the evolution of Galileo's mature thought, even for readers unwilling to consult Italian sources.

Drake's translation marks an important advance over the Crew De-Salvio version, placing the whole work more in its historical context and eliminating from it many unacceptable modernisations. In short, Drake's version is much closer to the original meaning, transferring many of Galileo's ambiguities and imprecisions into English equivalents. Galileo did not have the same precise terminology as does twentieth-century physics, though the Crew-De Salvio version often made it seem so. For some tastes, even Drake's more historically accurate version has perhaps not gone far enough in eliminating modernisms. Particularly some of the translator's own interpretative material leaves me a bit uneasy, *e.g.* the comparisons of Galileo to Hertz (notes on pp. 153, 159, 164) and Einstein (p. xxvii). Though Drake points out the tendency of his predecessors to make Galileo more 'experimental' than he was, he is not entirely able to escape this pitfall himself. It is hardly the translator's fault, however, for it is all but impossible to transfer the ambiguities of Latin and Italian terms such as *experientia, experimentum, comperio, esperienza, etc.* into fully equivalent English expressions, especially since there is no general agreement as to their precise meaning in the original. This gives the translator considerable latitude to impose an interpretation on the text. To cite but one instance, we find that in a single paragraph [critical edition, p. 122; Crew–De Salvio, p. 78; Drake, p. 81] the Italian term ***esperienza*** is used three times. Crew and De Salvio translate it consistently as 'experiment', making Galileo's discussion all sound very modern and carefully controlled. Drake on the other hand, translates the term once as 'experiment' and twice as 'experience', but gives no reason for translating the same word in two different ways. It seems clear to me that in none of the uses there is anything very close to what is conveyed by the modern English 'experiment', but still there is a little more (in this instance at least) meant than the vague 'experience'. Drake certainly handles the passage more skilfully than did his predecessors, but his rendering is still not wholly in conformity to the Italian, and he certainly realises the grave problems encountered by any conscientious translator. In brief, Drake does a very good job with what is basically an insoluble problem.

What this example points up is that, for any work of major importance which has been historically influential, a translation cannot satisfactorily be substituted for the original. At best translations can be devices to encourage students to recognise that there is a vast, significant, and challenging literature in foreign languages, but what should never be claimed is that translations are adequate substitutes for the original. Nevertheless, translations do have a value, even if it is more limited than generally admitted, at least insofar as historical or philosophical interpretation is concerned. Drake's version of *Two New Sciences* is probably as reliable as a translation can be expected to be, it reads easily and clearly and should remain standard for some time to come.

CHARLES B. SCHMITT
The Warburg Institute,
University of London

REFERENCE

WISAN, W. L. [1974]: 'The New Science of Motion: A Study of Galileo's *De motu locali*', *Archive for the History of the Exact Sciences*, 13, pp. 103–306.

MELTZER, B. and MICHIE, D. (*eds.*) [1972]: *Machine Intelligence 7*. Edinburgh: Edinburgh University Press. £12.00. Pp. xiv+485.

SCHANK, R. C. and COLBY, K. M. (*eds.*) [1974]: *Computer Models of Thought and Language*. San Francisco: W. H. Freeman. $13.50. Pp. vii+454.

These two books together provide a wide sampling of current work in artificial intelligence. *Computer Models of Thought and Language* (*CMTL*) concentrates on the relatively homely topics named in the title, while *Machine Intelligence 7* (*MI7*) ranges from these familiar phenomena, through heuristic search and robotics, to abstract questions of computational logic and program verification.

The former may be recommended to anyone concerned with human cognition, whether at a psychological or a philosophical level. The latter is more technical, and only a few chapters are of similarly general interest—the Preface, however, is startling, for in it Lord Bowden effectively suggests that no-one should waste any time in looking at either: 'Heaven knows these problems [Gestalt perception *etc.*] defy solution—why therefore should one expect that they can be studied or solved or practised by computers?'. He recommends AI workers to stop pursuing problems that are 'philosophically interesting rather than economically useful'.

This somewhat backhanded admission that the research in question is of philosophical interest may be wholeheartedly endorsed. For instance, *MI7*—the volume bearing the subversive Preface—includes an account by Winston of the *MIT* robot. This is a hand-eye system that can interpret a 2*D* photograph in terms of the original 3*D* scene of building blocks, and can use its inner representation of the scene in assembling a copy of the original from a scattered collection of variously shaped blocks. There are many points of interest in this achievement, one example of which is the system's perception of shadows in the scene.

Early scene analysis programs were unable to cope with shadows, and the lighting had to be 'fixed' accordingly. But the program which enables the *MIT* robot to see objects in shadowed scenes does not merely identify shadows so as then to ignore them: it uses them in its perception of the scene as a whole, so that shadows function as cues to matters of more substantive import. Moreover, this interpretative use of shadows suggests how the 'combinatorial explosion' may be effectively muffled: provided that it is used in an intelligent fashion, extra knowledge may speed thought up, rather than slowing it down.

The basic strategy of the object-recognition program is to label each vertex in the picture in terms of the physical properties of the edges and surfaces depicted: thus a line may represent a concave or a convex edge, and a surface may be occluding or not. If one allows picture-lines to depict shadow boundaries of various types as well as actual edges, the possible interpretations of any given line are increased. The alternative labellings of a given vertex (which of course contains several lines) are increased still more, to an apparently unmanageable number. However, it turns out that 'semantic' constraints on neighbouring label-sets considerably narrow the range of actual possibilities.

For instance, considerations of 3*D* geometry and the physics of illumination show that a common vertex which could in principle be labelled in 3×10^8 combinatorially different ways, in practice can represent only about 30 physical phenomena, so only these few internally consistent sets of labels need ever be seriously considered by the program: the semantics of the picture effectively keeps its syntax in check. What is more, sets of labels have to be ***mutually*** consistent as well as internally so. That is, a line cannot be labelled concave at one end and convex at the other if it is to be interpreted as representing a straight edge in the physical world; nor can it consistently be labelled both as an edge and as a shadow boundary. This allows for a 'jigsaw' phenomenon in the program's perception of the scene, without which it would be impossibly bogged down in endless computation. Even considering only the physically possible labellings of vertices, two adjacent vertices may each have as many as 500 possible interpretations independently. But since any line joining the two vertices must be labelled in the same way at either end, many of these can be immediately rejected by the program because they fail to fit properly together. This 'jigsaw' effect propagates constraints throughout the picture so that, amazingly, it often happens that only *one* interpretation of each vertex is consistent with its neighbours in context. In cases where some vertices do have more than one labelling left as a possibility after the 'consistency filtering', the program can carry out a systematic depth-first search to find which is the correct 3*D* interpretation of the 2*D* picture as a whole. The local ambiguity has already been precisely located (and its nature noted) *before* the systematic search begins Pictures that are perceived as ambiguous overall by people are typically interpreted by the program as being ambiguous in the same way.

Even these few remarks about contextual constraints and ambiguity suggest that—*pace* Lord Bowden—such work is not irrelevant to problems in Gestalt perception. Winston's paper indicates various ways in which 'wholistic' factors and complex schemata embodying knowledge and expectations influence the perceptual interpretation of potentially ambiguous stimulus figures. Further, the discussion of the *building* done by the *MIT* robot when copying its perceived model raises issues concerning general problem solving and the planning of purposive action in the environment.

Sceptics who, like Lord Bowden, feel that the understanding of language is a problem which similarly 'defies solution' in computational terms should read the papers by Winograd (*CMTL* ch. 4), Schank (*CMTL* ch. 5), Wilks, (*CMTL* ch. 6), and Davies and Isard (*MI7* ch. 17). I have reviewed Winograd's remarkable program in an earlier issue of this *Journal*—it is essential reading for anyone seriously interested in language. But it suffers from being neither syntactically flexible nor semantically deep: it has to have a perfect sentence as input, it can talk only about simple building blocks, and it has virtually no ability to anticipate the line of conversation and make 'intelligent' contributions accordingly. Schank's work—which is largely programmatic rather than fully programmed—offers the possibility of overcoming these drawbacks.

Schank's basic insight is that the intelligent speaker, whether man or machine, is continually making predictions about the conceptual structure of that part of the message which has yet to be communicated. This 'effort after meaning' contributes essentially to fruitful communication, and is consciously brought to light only when the expectations are disappointed so that one wonders 'What has that got to do with it?'. Schank develops a semantic system of 'conceptual dependencies', wherein the myriad verbs of everyday language are to be analysed in terms of only fourteen primitive actions. These are related to each other by way of a small number of basic semantic elements, such as causation, tense, attribution, and the four 'conceptual cases'—objective, recipient, directive, and instrumental. These cases are crucial in setting up the expectations that structure one's interpretation of other people's utterances, and are exploited by Schankian programs in understanding and responding to input remarks.

Wilks also describes a semantic theory that has been partially implemented in computational form, but his main interest is in machine translation. In terms of his semantic primitives he lists a finite number of messages, or 'gists', which supposedly underlie all intelligible communications and so limit word-substitutions during translation. The objection that language can be used creatively (analogically or metaphorically), so that no communication can be ruled out in principle, is countered by Wilks's theory of 'semantic preferences'. 'Drink', for instance, prefers but does not demand an animate subject: Wilks's program is competent to render 'The car drank the petrol' into French while recognising the atypical use of the verb. Finally, Davies and Isard offer some interesting remarks about various types of speech-act in their comparison of utterances with programs. Light is thrown, for instance, on the Russell–Strawson controversy about sentences referring to 'the present King of France': the authors suggest that the rejection of a sentence on the grounds of 'false presuppositions' is best represented by the failure of some bit of program which the hearer attempts to run during the compiling of the main program specified by the speaker.

The phenomenon of language is intimately related to that of belief. Indeed, the computer models of belief systems described by Abelson (*CMTL* ch. 7) and Colby (*CMTL* ch. 6) are closely bound up with computer representations of language. Thus Abelson sketches a six-level theory of beliefs, the three basic levels of which are heavily dependent on Schank's semantic system. Abelson's main interests are in the psychology of political ideology and attitude change, and he has already implemented several relevant computer simulations. His recent work systematises the abstract cognitive structures with which one intuitively characterises interpersonal realities such as victory, alliance, love, betrayal, revolution—and even the end of the honeymoon. Such matters are represented in terms of the interdependent plans of the two actors (individuals or groups) concerned, each of which may have various sorts of role in the other's plan. They may be agents for one another or be interested parties in each other's success, they may have evaluative attitudes to each other's plans, and they may be able to help or hinder each other in different ways. Abelson draws up an explicit taxonomy of belief structures which he relates to his earlier work on the simulation of political conservatism, and which could be used as a theoretical guide directing conceptual or empirical investigations of and comparisons between indefinitely many individuals or cultures.

Colby's work also allows for comparisons between different belief systems, but his prime concern is the distinction between normal, neurotic, and psychotic (specifically: paranoid) beliefs. Someone who is paranoid about a certain topic responds to remarks about it differently from someone who is not. The response need not be explicit: though he keeps silent, he may inwardly interpret the remarks as hostile and become increasingly angry with his interlocutor accordingly. Analogously, neurotics inwardly transform anxiety-ridden beliefs about personal relationships into other, less threatening, beliefs. But all people, whether 'normal' or not, structure their beliefs and respond to conversation by way of a central conceptual system which can be fruitfully compared with others in various ways. Colby's artificial neurotic, for instance, transforms affectively loaded beliefs in the sorts of ways characterised by Freud as 'defensive'. And his recent 'Paranoid' program incorporates a number of evaluatively laden 'core' concepts comprising a delusion of persecution by the Mafia, together with a theory of paranoid thought processes which directs the conceptual transformations carried out during conversations with the human operator. Despite the many limitations of Colby's programs, psychiatrists asked to diagnose a patient on the basis of teletype interviews cannot distinguish between human and artificial paranoiacs. More important, computer models such as these are useful in enabling one to raise relatively clear questions about the underlying cognitive processes involved in human thought.

This clarificatory advantage of computer implementations is stressed by Newell, in a paper relating artificial intelligence to the concept of mind (*CMTL* ch. 1). As well as discussing the general psychological relevance of machine research, he describes some of his own work on problem solving in cryptarithmetic. According to Newell, problem solving is best represented in terms of sets of rules termed 'production systems', which specify transformations of one expression into another, given certain conditions. A rather different approach to problem solving is found in a paper by Fikes, Hart, and Nilsson, which describes the thinking capacity of the Stanford robot (*MI7* ch. 23). They ask how the structure of a complex problem may be represented so as to enable the robot problem solver to formulate an overall plan and translate strategy into tactics in a fruitful and flexible manner. They also discuss how a specific solution may be generalised, and stored by the program for use in other circumstances. Their account raises a number of questions concerning the structure of purposive action, including how one represents 'negative' goals and unwanted side-effects of *prima facie* appropriate actions.

In this review I have concentrated on the papers of psychological relevance, and have not even mentioned all of these. Other papers, concerning scientific hypothesis formation (*MI7* ch. 14) and more strictly logical issues (*MI7* chs. 2–12) should also be of interest to philosophers of science. In sum, whether or not these books report work that is 'economically useful', they unquestionably contain matter which is 'philosophically interesting'.

MARGARET A. BODEN
University of Sussex

EIGHTH INTERNATIONAL CONGRESS ON CYBERNETICS

The International Association for Cybernetics is organising the Eighth International Congress on Cybernetics. It will take place in Namur (Belgium) from 6th to 11th September 1976.

Subjects of the Congress:

1. Principles of Cybernetics and General Theory of Systems;
2. Cybernetics in Social Systems;
3. Cybernetics in Mechanical Systems;
4. Cybernetics in Biology and Medicine.

All those who are interested by the subjects of the Congress may take part.

Authors of papers are requested to send the title and a summary of the papers before 1 January 1976.

Information and documentation about the Congress are available at the Secretariat of the International Association for Cybernetics, Palais des Expositions, Place André Rijckmans, B-5000 Namur (Belgium).

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER

SEPTEMBER, 1975, VOLUME XXXVI, No. 1

CONTENTS

I. ARTICLES
- The Information-Processing Approach to the Brain-Mind and its Philosophical Ramifications, by C. A. HOOKER
- Towards a Phenomenological Mathematics, by J. G. FAUVEL
- Self-Esteem and Human Equality, by LYNNE BELAIEF
- Fallibility and Knowledge of the Future, by JOHN FISHER
- On Believing Inscriptions to be True, by MICHAEL CORRADO

II. DISCUSSIONS

III. REVIEWS

IV. NOTES AND NEWS, RECENT PUBLICATIONS

Subscription prices $8.00 per year for individuals; $10.00 per year for institutions. Separate issues, $2.00 and $2.50 respectively. Postage and handling for back issues only. Papers submitted for publication will not be returned unless accompanied by a self-addressed stamped envelope, or equivalent postage.

A Circular listing the main contents of the Journal since 1940 will be sent upon request.

Address all communications to

PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH

State University of New York at Buffalo, Buffalo, New York, 14226

Philosophy of Science

OFFICIAL JOURNAL OF THE PHILOSOPHY OF SCIENCE ASSOCIATION

Editorial Board

Richard S. Rudner (Editor-in-Chief), Peter Achinstein, Gustav Bergmann, May Brodbeck, Arthur Burks, C. West Churchman, Theodosius Dobzhansky, Adolf Grünbaum, Edward Madden, Henry Margenau, Gerald J. Massey, Alberto Pasquinelli, Wesley C. Salmon, Kenneth F. Schaffner, Dudley Shapere, J. J. C. Smart, Melford E. Spiro, Patrick C. Suppes, Lewis K. Zerby, Peter D. Asquith (Managing Editor). Book Reviews, James Van Evra. Assistant Editors, Patricia Ann Fleming and Paul C. L. Tang.

CONTENTS*

VOLUME 42 (1975) will include

DEBORAH ROSEN. *An Argument for the Logical Notion of a Memory Trace*

RONALD C. HOY. *The Role of Genidenity in the Causal Theory of Time*

HERBERT HENDRY. *Ramsey Sentences for Infinite Theories*

JEAN-CLAUDE FALMAGNE. *A Set of Independent Axioms for Positive Hölder Systems*

C. A. HOOKER. *On Global Theories*

S. WENDELL-WAECHTLER and E. LEVY. *More Philosophical Aspects of Molecular Biology*

ARTHUR E. FALK. *Learning to Report One's Introspections*

DAVID G. BLAIR. *On Purely Probabilistic Theories of Scientific Inference*

BROWN GRIER. *Prediction, Explanation and Testability as Criteria for Judging Statistical Theories*

*The contents of *Philosophy of Science* are indexed in *Science Citation Index, ABC Pol Sci, Language and Language Behaviour Abstracts,* and *The Philosopher's Index.*

PHILOSOPHY OF SCIENCE is published quarterly by the Philosophy of Science Association, 18 Morrill Hall, Department of Philosophy, Michigan State University, East Lansing, Michigan 48824. Subscription price $17.50 ($18.50 non-U.S.A.) per Volume. Members of the Association receive the Journal as a privilege of membership paid for out of membership dues. Inquiries about subscription or membership should be directed to the Managing Editor.

MIND

A Quarterly Review of Philosophy

edited by Professor **DAVID W. HAMLYN**

Vol. LXXXIV October 1975 No. 336

Articles:

Reasons for Action, Decisions and Norms	J. RAZ
Blame and Causality	L. HERTZBERG
Timelessness and Foreknowledge	P. HELM
Games and Language	C. CHERRY
Heidegger and Death as 'Possibility'	P. EDWARDS
Description and Depiction	E. M. ZEMACH

Discussions:

Blumberg on Moral Criticism	L. V. BRETTLER
When Do Causes Take Effect?	I. THALBERG
Solitary Speakers	P. SMITH
Essence and Accident	M. DURRANT
Frege and Wittgenstein on Elucidations	P. M. S. HACKER
A Retort to the Tortoise	M. HOLLIS

Book Reviews

Books Received

Annual Subscription £2.50, single issues 80p

Published for the MIND ASSOCIATION by

BASIL BLACKWELL **49 BROAD STREET OXFORD**

RATIO

edited by Stephan Körner

VOLUME XVII DECEMBER 1975 No. 2

FREDERIC B. FITCH	Induction and the Existence of God
R. BRECHER	Hartshorne's Modal Argument for the Existence of God
KNUT ERIK TRANØY	'Ought' Implies 'Can': A Bridge from Fact to Norm? (Part II)
REINHARD BRANDT	Historical Observations on the Genesis of the Three-Dimensional Optical Picture
TERRENCE F. ACKERMAN	The Literal, the Metaphorical, and the Real
R. A. SHARPE	On the Causal Theory of Knowledge
NORMAN SWARTZ	Spatial Worlds and Temporal Worlds: Could there be More than one of Each
BRIAN SMART	The Statues and a Problem of Identity
ANTHONY C. GENOVA	Opacity, Inexistence and Intentionality
ALBERT SWEET	A Pragmatic Model of the Epistemic Logic of Chisholm and Keim
REVIEWS:	
D. H. MELLOR	*The Cement of the Universe: a Study of Causation*, by J. L. Mackie
JONATHAN HARRISON	*The Development of Bertrand Russell's Philosophy*, by Ronald Jager
ROSS HARRISON	*Human Understanding, Volume I: The Collective Use and Evolution of Concepts*, by Stephen Toulmin
NOTE	
EDDY M. ZEMACH	On Wittgenstein's *Sachverhalte*
BOOKS RECEIVED	

Price £3.25 net **Annual Subscription: £6.00 net**

Orders to: Journals Department, Basil Blackwell & Mott Ltd., 108 Cowley Road, Oxford OX4 1JF, England

Published by

BASIL BLACKWELL · OXFORD

Revue Internationale De Philosophie

Direction scientifique: Professeurs Ph. DEVAUX, de l'Académie royale de Belgique
H. GOUHIER, de l'Institut de France
Ch. PERELMAN, de l'Académie royale de Belgique
A. ROBINET, du C.N.R.S. de France & de l'Université de Bruxelles
Direction administrative, Rédaction: Mr. M. SERVAIS
99 avenue de l'Université-1050 Bruxelles-Belgique
Tel. 02/648 03 70

No. 109—1974—Fasc. 3

PROBLEMES ACTUELS DE L'ESTHETIQUE

Sommaire:

E. Souriau, *L'art comme symptôme philosophique*
W. Tatarkiewicz, *Notes sur la perfection de l'art*
H. Osborne, *Aesthetics and Value*
M. Dufrenne, *Création et engagement politique*
A. Souriau, *Esthétique et ontologie*
P. Minguet, *Sur l'esthétique du désoeuvrement*
M. C. Beardsley, *The Descriptivist account of Aesthetic Attributions*
G. Vattimo, *Arte e identità-Sull'attualita dell'estetica di Nietzsche*

Variete:

R. Legros, *L'égalité dans la pensée politique de l'Utilitarisme*

Analyse:

P. Gochet, *Esquisse d'une théorie nominaliste de la proposition-Essai sur la philosophie de la logique.* (J. Ruytinx)

OUVRAGES RECUS—REVUE DES REVUES

No. 110—1974—Fasc. 4

CASSIRER

Sommaire:

S. K. Langer, '*De Profundis*'
H. Blumenberg, *Ernst Cassirers Gedenkend*
N. Rotenstreich, *Schematism and Freedom*
J. E. Smith, *Some Comments on Cassirer's Interpretation of Religion*
R. Nadeau, *Bibliographie des textes sur Ernst Cassirer*

Varietes:

J. Kupfer, *Universalization in Berkeley's Rule-Utilitarianism*
A. Robinet, *Descartes, Malebranche et Ferdinand Alquié*

Analyse:

Studi sull'Uguaglianza—(E. Griffin-Collart)

OUVRAGES RECUS—REVUE DES REVUES—VIE PHILOSOPHIQUE—TABLE DES MATIERES DU TOME XXVIII (1974)

Abonnements:
Belgique: 530 Fb; Pays Européens: 580 Fb; Autres pays: 620 Fb.

INQUIRY

An Interdisciplinary Journal of
Philosophy and the Social Sciences
Edited by Arne Naess
and Alastair Hannay

Number 3 *Autumn 1975* *Volume 18*

Some Myths about 'Mental Illness' — Michael S. Moore
Wittgenstein on Ostensive Definition — P. M. S. Hacker
Doubts about Moral Principles — Peter Mew
Bill Wallace: A Conversation — A. Dewey Jensen
Towards a Spinozistic Modification of Skinner's Theory of Man — Carl G. Hedman
Discussion:
What is the Ethical in *Fear and Trembling*? — Bruce Russell
Review Discussions:
Galileo and the Scientific Revolution (D. Shapere: *Galileo. A Philosophical Study*) — Per Strømholm
A Kind of Philosopher: Some Recent Books on Kierkegaard — Alastair Hannay

Published quarterly by: Universitetsforlaget, P.O. Box 307
Blindern, Oslo 3, *Norway*

Annual Subscription: Institutions—N. kr. 75.00, £4.90, $15.00
Individuals—N. kr. 60.00, £3.90, $15.00
Postage included

Philosophical Studies

An International Journal for Philosophy in the Analytic Tradition

Editor: Keith Lehrer, University of Rochester, Rochester, U.S.A.

Board of Consulting Editors: David Braybrooke, Joseph L. Camp, Jr., L. J. Cohen, James Cornman, Alan Donagan, Keith Donnellan, Mary Hesse, Risto Hilpinen, R. M. Martin, Terence Parsons, Günther Patzig, A. M. Quinton, Jerome Schneewind, Robert Sleigh, Jr., Judith Thomson.

Philosophical Studies fills the urgent and growing need for space devoted to quick publication of timely contributions to analytic philosophy. It emphasizes the rapid exchange of notes and criticism. To provide balance and scope the editors will, on occasion, solicit contributions, comments and elucidations from philosophers whose voices might not otherwise be heard when the spotlight of discussion points in their direction.

Philosophical Studies welcomes papers in all areas of analytic philosophy. It recognizes that there are divergent 'tendencies' within the analytic movement, and requires only that contributions represent an application of methods of analysis and clarification of meaning. Papers the content of which falls primarily within the scope of the natural, social or psychological disciplines will be considered only if this content is subordinated to the purposes of philosophical analysis.

Subscription price per volume of six issues Dfl. 90,- / US $ 36.00 plus Dfl. 17,50 / US $ 7.00 for postage and handling. Nine issues will be published yearly. Personal subscription price on request.

D. Reidel Publishing Company
Dordrecht - Holland / Boston - U.S.A.

12 DEC 1975

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1975-6

President: Professor D. J. O'Connor
Vice-President: Professor J. W. N. Watkins
Hon. Secretary: Helena Sheiham
Hon. Treasurer: Mr. D. Miller
Members: Mr. C. Howson, Dr M. Machover, Mr H. Martins, Mr. A. Young

EDITORIAL BOARD OF THE JOURNAL

Professor R. B. Braithwaite
Professor H. Dingle
Professor Sir John Eccles
Professor Sir Harold Jeffreys
Professor W. C. Kneale
Professor S. Körner
Professor M. Polanyi
Professor Sir Karl Popper
Professor G. J. Whitrow
Professor J. H. Woodger

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules :

' The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all persons who have been proposed by two members, approved by the Committee, and elected at an ordinary meeting. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £1·00 ($2.50). Those who wish to join should write to the Hon. Secretary, Helena Sheiham, Department of Philosophy, London School of Economics, Houghton Street, London WC2A 2AE.

Brit. J. Phil. Sci. Vol. 26 No. 3 September 1975

The British Journal for the Philosophy of Science

Contents

Articles

187–198 COLIN HOWSON The Rule of Succession, Inductive Logic and Probability Logic
199–211 PAUL HORWICH Grünbaum on the Metric of Space and Time
213–225 YORICK WILKS Putnam and Clarke and Mind and Body

Discussions

227–232 ARTHUR THOMAS Puccetti on Machine Pattern Recognition
232–234 J. W. ROXBEE COX Mackie on Dispositional Properties
234–239 JAMES CARGILE Newcomb's Paradox

Review Articles

241–253 L. PEARCE WILLIAMS Should Philosophers be Allowed to Write History? (Review of J. Agassi: *Faraday as Natural Philosopher* and W. Berkson: *Forces and Fields*)
255–264 MAURICE FINOCCHIARO Philosophizing about Galileo (Review of D. Shapere: *Galileo: A Philosophical Study*)

Reviews

265–273 RUPERT HALL on N. Bukharin (*ed.*): *Science at the Cross Roads*; CHARLES SCHMITT on Galileo's: *Two New Sciences*, translated by Stillman Drake; MARGARET BODEN on B. Meltzer and D. Michie (*eds.*): *Machine Intelligence* 7 and R. C. Schank and K. M. Colby (*eds.*): *Computer Models of Thought and Language.*

Announcement

12 DEC 1975

Printed in Great Britain at the University Press, Aberdeen

Volume 26 Number 4 December 1975
BJPIA5 26 (No. 4) 275-363 (1975) ISSN 0007-0882

The British Journal for the Philosophy of Science

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editors: JOHN WATKINS and JOHN WORRALL

Assistant Editors: COLIN HOWSON and ELIE ZAHAR

Subscriptions

The *Journal* is published quarterly. The subscription price is £7.00 net (U.S.A. and Canada U.S. $22.00) per volume post free; single issues £2.50 net (U.S.A. and Canada U.S. $8.00) plus postage. The subscription is £4.50 (U.S.A. and Canada U.S. $14.00) for members of the British Society for the Philosophy of Science, who should order direct from the publisher. (Membership of the Society is international and is by no means confined to people resident in the United Kingdom.) Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

Notes to Contributors

Articles, correspondence and books for review should be sent to John Worrall at the Department of Philosophy, London School of Economics, Houghton Street, Aldwych, London WC2A 2AE. Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Footnotes, which should be as few as possible, should be typed separately, double-spaced at the end of the text. They should be referred to by consecutive numbering throughout the text, in which they will be inserted in page proof. Bibliographical references should be listed alphabetically at the end. The style of references should follow current *Journal* practice.

Long quotations are printed in small type without quotes, and should be so marked in the typescript. Single quotes should be used to form names of words and for shorter quotations; double quotes for other purposes. Diagrams, tables, and illustrations should be on separate sheets, with their desired position in the text clearly indicated. Symbolic formulae appearing as separate lines should be clearly distinguished in the text, and foreign letters (Greek, German, etc.) indicated in the margin. Material to be printed in italics, especially English letters used as symbols ('*A*', '*x*', etc.), should be underlined. On matters of style and layout, authors should consult CHAUNDY, T. W. *et al.* [1964]: *The Printing of Mathematics*, London: Oxford University Press.

Offprints of articles may be ordered when corrected galley-proofs are returned. Fifty are supplied free; the price of extra copies is given on the order form.

© British Society for the Philosophy of Science 1975

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editors.

Back Volumes

Volumes 1-16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17-25 may be obtained from Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Prices on application.

Brit. J. Phil. Sci. 26 (1975), 275–301 *Printed in Great Britain*

Antipositivism in Contemporary Philosophy of Social Science and Humanities

by JERZY GIEDYMIN

1 *Positivism in the Strict Sense and 64 Mutations.*
2 *Positivism as Phenomenalism and Instrumentalism.*
3 *Positivism as Naturalism.*
4 *Positivism and Positive Economics.*

The dominant tendency in contemporary philosophy of natural and social science is antipositivist. That much is easy to ascertain even by a cursory survey of books and articles. But what are antipositivists against? Whether dead or alive, their enemy is difficult to identify on the basis of antipositivist accounts which vary considerably one from another. A brief glance at those accounts will suffice to show that antipositivists do not have an identity either, they are a motley group with different interests, motives and backgrounds.

In the present article I shall restrict my attention to those antipositivists who are, broadly speaking, in the empiricist tradition and whose main complaint against positivism is that as a descriptive account of empirical science it is too narrow and as a methodological programme it is a too restrictive brand of empiricism. I shall argue that the basic antipositivist criticisms are relevant and effective against what I shall call positivism in the strict sense but not against a variety of scientists and philosophers also labelled positivists in some vague sense of the word. It is, therefore, in the antipositivists' own interest—unless their aim is confusion rather than rational discussion—to identify their opponents more precisely. I shall also argue that whereas positivism in the strict sense is in fact for many purposes an unnecessarily restrictive brand of empiricism, this does not apply to a variety of doctrines or programmes often subsumed under 'positivism' in the broad and vague sense and that many views and principles the antipositivists regard as their own (and so non-positivist) have been part of their opponents' doctrines for some time now.

there is a common ground

Received 1 *July* 1975

T

I POSITIVISM IN THE STRICT SENSE AND 64 MUTATIONS

By positivism in the strict sense I shall mean here the positivism of August Comte and of John Stuart Mill, as well as the positivism of the Vienna Circle, *i.e.* so-called logical positivism of Schlick, Neurath, Carnap and others, as developed in the late nineteen-twenties and early nineteen-thirties (before 1935). Since I want to take into account not all views of Comte (*e.g.* his views on science but not his 'religion of humanity') and since there are some differences between the philosophy of Comte and Mill, not to mention logical positivists, it is desirable to give an explicit list of doctrines to be labelled 'strict positivism'. The list comprises the following items:

(*a*) identification of knowledge with science (natural and social) and mathematics, to the exclusion of other areas, *e.g.* ethics;

(*b*) empiricism in the extreme form of either phenomenalism or physicalism, *i.e.* the reduction of science to statements about directly observable facts and the elimination as meaningless of any sentence that is neither analytic nor empirical, *e.g.* of metaphysics;

(*c*) the reduction of philosophy to the 'logic of science' (philosophy of science) and of mathematics;

(*d*) methodological naturalism (naturalistic methodological monism), *i.e.* the view that the social sciences and even humanities have basically the same aims and methods as the natural sciences;

(*e*) sociological relativism with respect to norms, in particular ethical ones;

(*f*) the emphasis on the social value of science and on its practical applications.

These doctrines may be formulated either as descriptive statements (with the possible exception of (*f*)) or as rules. Since each may be affirmed or denied (be present unaltered on the list or not) independently of others and since they are six in number, there are $2^6 = 64$ possible lists with varying positivist content. Just how many and which of these doctrines (or programmatic rules) have to be on an author's list in order to regard him as a positivist (antipositivist) is an arbitrary terminological decision. Unless such a decision is made, for example by identifying positivism with strict positivism (the complete list (*a*) through (*f*)), the concepts of positivism and antipositivism are vague. To use them either as if they were not vague or to imply that the acceptance of one or of only some of the doctrines commits one logically to the acceptance of the remaining ones, is to produce confusions.

In the non-strict or wider sense one refers to any philosophy as positivist or as containing positivist elements if it strongly emphasises the anti-speculative attitude in both scientific theorising and in philosophy, the ideals of caution, clarity and precision, the preference for scientifically solvable and practically useful problems. The first two are clearly related to (*a*), (*b*) and (*c*), the third to (*f*). So, for example, in ancient Greece the distinction made between mathematical and physical astronomy may be said to have helped practise mathematical astronomy in a somewhat positivist spirit; for it restricted the job of a mathematical astronomer to the invention of hypotheses or 'expedients' by the assumption of which the observed phenomena would be 'saved' (computed or predicted) and excused him from engaging in causal explanations in terms of Aristotelian physics and ontology; but this did not amount, of course, to strict positivist philosophy, since most ancient astronomers would reject (*a*), (*b*) and (*c*). Similarly, when Osiander in his uninvited preface to Copernicus's *De Revolutionibus* wrote that the Copernican 'hypotheses need not be true nor even probable; if they provide a calculus consistent with the observations, that alone is sufficient', he may be said to have suggested a somewhat positivist view of the heliocentric theory but he certainly was not a positivist since he believed in divine revelation as a source of knowledge thus rejecting at least (*a*) and (*c*). The methodological principle known as 'Occam's razor' which advises 'not to multiply entities (for accounting for the phenomena) beyond necessity', is usually seen as an expression of the somewhat positivist anti-speculative attitude associated with Occam's nominalism in logic and as a forerunner of (*b*). But Occam did not repudiate metaphysics or theology although he insisted that logic and science could and should be studied without reference to either metaphysics or theology. Berkeley's theory of meaning, on which his criticism of the concept of substance and of Newtonian concepts of absolute time and motion were based, was definitely one of the forerunners of (*b*), as was Hume's critical analysis of the concept of causal relation (causal law). Hume's celebrated condemnation of any treatise that did not contain 'matters of fact' or 'relations between ideas' was obviously a *locus classicus* of the anti-metaphysical doctrine of positivism, but neither Berkeley nor Hume were positivists in the strict sense of (*a*) through (*f*) as both rejected at least (*c*) and (*e*).

Two of the doctrines of strict positivism, namely (*b*) and (*d*), have played a prominent role in extending the application of the concept of positivism and require special attention not only on that account alone but also because, in my view, they have given rise to *confused* or *misguided anti-positivism.* This applies to the philosophy of science in general but particularly to the philosophy of the social sciences and humanities.

Positivism restricted to doctrine (*b*) amounts to *phenomenalism* or *physicalism*. Both appear either as descriptive accounts of science or as programmes how to practise science. Both contain criteria of empirical meaningfulness of terms and sentences and claims concerning the status of theoretical terms and sentences. As doctrines concerning the cognitive status of theories they are also known under the name of '*descriptivism*' and '*instrumentalism*'.

Positivism restricted to doctrine (*d*) is, of course, identical with *methodological naturalism* or naturalistic methodological monism, and as such is most often discussed and criticised in contemporary antipositivist literature in the philosophy of the social sciences and humanities.

We shall see that the controversy between methodological naturalism and its antipositivist opponent, methodological antinaturalism or dualism, is not independent of one's attitude towards phenomenalism, physicalism and instrumentalism, both as descriptive accounts of science and as programmes. I shall discuss, therefore, first positivism in the sense of phenomenalism, physicalism and instrumentalism and then the antipositivist criticisms of methodological naturalism.

2 POSITIVISM AS PHENOMENALISM AND INSTRUMENTALISM

One of the leading critics of strict positivism in contemporary philosophy of natural science, Sir Karl Popper, is also the author of one of the fiercest criticisms of instrumentalism. I shall now briefly report and comment on Popper's views on instrumentalism, emphasising especially those points which are relevant to the issue of positivism as methodological naturalism.

Popper contrasted three views concerning the status of physical theories, essentialism, instrumentalism and the third view, *i.e.* realism combined with fallibilism or hypotheticism, which is also Popper's own view.[1] I shall disregard here essentialism and concentrate on the contrast between instrumentalism, as understood by Popper, and his 'third view'.

Instrumentalism was identified by Popper with the view concerning the status of scientific theories '. . . founded by Osiander, Cardinal Bellarmino and Bishop Berkeley . . .' but also advocated in various ways by Mach, Kirchhoff, Hertz, Duhem, Poincaré, Bridgman, Eddington, as well as Niels Bohr and the so-called Copenhagen school. The following (summarised) tenets are, according to Popper, characteristic of instrumentalism:

> Scientific laws and theories are not proper descriptive statements; they are nothing but instruments to derive observational predictions from other observational statements. Dispositional and theoretical terms (*e.g.* magnetic, force,

---

[1] Popper [1963].

field of forces, *etc.*), unlike observational terms, do not refer to real entities but are merely symbols which facilitate deductions. When a theory is found not to 'work' in a given area, its field of application is suitably modified. Since scientific theories are nothing but computational rules of the same character as the computational rules of 'applied sciences', there is no difference between 'pure' and 'applied' science; all science is applied; in particular, explanation is not among the aims of science, only description and prediction.

By contrast, according to Popper's 'third view' (realism-*cum* hypotheticism), scientific laws and theories are not only instruments for making predictions but also descriptive and explanatory hypotheses. Dispositional and theoretical terms are intended to refer to real entities just as observational terms; in any case, all terms are theoretical to some degree. Scientific hypotheses and theories, though not verifiable by observation, can clash with reality, provided they are falsifiable; experiments and observations, though unable to establish any theory, can eliminate some as false. The fundamental aim of science is to provide a (hopefully) true description of the world, explanations of increasing depth and predictions, some of which at least are of a new kind, *i.e.* are real discoveries.[1]

There are several reasons why, according to Popper, instrumentalism gained popularity among scientists and philosophers. One of those reasons was the fact that it enabled them 'to deal with inconvenient scientific hypotheses'. So, for example, it enabled Osiander and Bellarmino to deal with the Copernican hypothesis in such a way that it no longer appeared to contradict the Scripture. Another reason, related to the previous one, was the concern that if science is credited with the ability to discover the truth about the world unaided by divine revelation, religion would suffer. This was, for example, Berkeley's concern. There was a third reason. Many scientists and philosophers rejected the essentialist view that the aim of science was to discover hidden causes or essences of things, either because they did not believe that such hidden causes or essences existed (*e.g.* Mach) or because they did not believe science was able with its methods to discover them (*e.g.* Berkeley, Duhem, Poincaré). Consequently they denied that theoretical concepts referred to latent causes or properties and that the aim of science was to give explanations which transcend our knowledge of the phenomenal world. Similarly, from the correct belief that science is unable to demonstrate the truth of its theories and that so-called crucial experiments cannot establish any theory as true, instrumentalists (*e.g.* Osiander, Bellarmino, Duhem, Poincaré) drew the wrong conclusion that scientific theories are never descriptive, possibly true statements but are merely mathematical tools for making predictions.

---

[1] *Ibid.*, 'Theories as instruments', pp. 107–11; 'The third view', pp. 114–19.

Similarly, from the (correct) premise that the method of science cannot yield but tentative, revisable hypotheses and from the further premise that scientific laws and theories (*e.g.* Newton's laws of motion, conservation laws) cannot be contradicted by experimental findings nor can a crucial experiment be instituted to refute, *e.g.* one (physical) geometry in competition with another, some instrumentalists (*e.g.* Poincaré) drew the wrong conclusion that scientific laws and theories are nothing but terminological conventions, useful for systematising observational data (in the way a catalogue systematises books in a library), but having no descriptive, empirical content.

In his evaluation of instrumentalism Popper claims that as a descriptive account of science instrumentalism is inadequate and false. In its effect on science it is, moreover, obscurantist, it hampers the progress of science. Similar views were voiced by others, *e.g.* P. Feyerabend.

On Popper's account, instrumentalism may be understood as resulting from phenomenalistic (physicalistic, operationalist) criteria of meaning accepted either explicitly or implicitly by instrumentalists. According to such criteria only observational expressions, *i.e.* observational terms, terms definable explicitly by observables and observational sentences are (empirically) meaningful. A theory is, on these criteria, always logically equivalent to a set of observational (phenomenalistic) sentences. (This view of theories is also known as descriptivism). If a theory contains terms which are not explicitly definable by observables, then the sentences in which such terms occur—so-called theoretical sentences—are not (empirically) meaningful; they are either analytic (terminological conventions) or else purely formal symbols—in either case they are non-descriptive.

In my opinion, Popper's account and criticism of instrumentalism have various shortcomings, which—however—need not be discussed here in detail.[1] For the purposes of the present article it will be sufficient to distinguish between instrumentalism in Popper's sense and—what I shall call—*moderate instrumentalism*. According to the latter, a scientific theory may contain, apart from observational sentences and empirical laws, theoretical sentences which are either empirical though not equivalent to any observational sentences, or analytic or else indeterminate (neither true nor false). Seen from another point of view, a theory may also contain some empirical, theoretical sentences which are not asserted at the given moment (they may be asserted at some later time or else their negations may be asserted). According to the moderate instrumentalist view, therefore, at least some theoretical sentences may be empirical and may function as premises in explanations of the nomological-deductive type; others are neither true nor false and yet serve as premises in 'explanations' conceived as re-descriptions of observable facts in the language of 'scientific metaphors' or 'fictions'. In fact, most authors mentioned by Popper as instrumentalists were, in my view, moderate instrumentalists in the sense just specified; this applies to Osiander, Bellarmino, Poincaré, Duhem and to some logical empiricists, *e.g.* Rudolf Carnap (after 1935); for Osiander and Bellarmino did not claim that the Copernican hypotheses were neither true nor false; Poincaré and Duhem did not claim that all scientific laws and theories were purely conventional and in Carnap's logical reconstruction theoretical sentences may be empirical.

---

[1] I have criticised Popper's account and critical analysis of instrumentalism in an article 'Instrumentalism and its Critique: A re-appraisal' to appear in a forthcoming volume of the *Boston Studies in the Philosophy of Science*.

It is also worth mentioning the fact, disregarded by Popper and other critics of instrumentalism, that the instrumentalist view of theories has been associated with what might be called the *poly-theoretic model* of scientific method. Descriptively, the poly-theoretic model represents the method of science as a (rational) choice of one, or *several* of a number of alternative hypotheses or theories on the basis of various choice criteria; from the normative viewpoint, it recommends the 'proliferation' of hypotheses or theories. The poly-theoretic model of scientific method may be found in the writings of ancient astronomers, in Osiander's preface and letters, as well as in the writings of conventionalists. By contrast, realism (not to be confused with Popper's 'third view') has been associated with a *mono-theoretic model* of scientific method, according to which scientists work (and ought to work) with one single theory at a time. A classical expression of the idea of the mono-theoretical model of science may be seen in Newton's Rule IV, which implied that '. . . the argument of induction may not be evaded by hypotheses'. Which of the two models of scientific method is descriptively and normatively better is perhaps a controversial question. However, the poly-theoretic model seems to be regarded as much more satisfactory by contemporary philosophers of science, including Popper and other critics of instrumentalism. For this and other reasons, it is difficult to agree with Popper's unqualified condemnation of instrumentalism.

If phenomenalism and instrumentalism are identified with positivism, then any realist view of scientific theories, *e.g.* Popper's realism-*cum*-hypotheticism, is non-positivist. However, since there are intermediate positions between strict realism and instrumentalism, *e.g.* moderate instrumentalism, the distinction between positivism and anti-positivism in this sense is no longer sharp and clear. Moreover, nineteenth-century moderate instrumentalism is related, through its neo-Kantian origin, to the 'hermeneutic' trend in the philosophy of humanities. Finally, those who are strongly influenced by later Wittgensteinian philosophy, *e.g.* S. Toulmin, T. Kuhn, are much closer to instrumentalism than to realism and yet are self-confessed anti-positivists.

3 POSITIVISM AS NATURALISM

Since the issues between the positivist and anti-positivist philosophy of natural science are so complicated when the realism-instrumentalism controversy is taken into account, it is not surprising that the dispute in the philosophy of social science between positivism and anti-positivism is even more involved when the former is identified with methodological naturalism (monism) and the latter with methodological antinaturalism. For the claim of naturalism that the natural and social sciences have fundamentally the same aims and methods is no longer unambiguous when those aims and methods are characterised differently by phenomenalists and their critics, by instrumentalists and realists; the interaction between the opposing doctrines is responsible for further diversification. One's methodological diagnosis of social science in relation to natural science depends, of course, as much on how one sees the latter as on how one sees the former. All these complicating factors are usually ignored in contemporary accounts of positivism and antipositivism. As a consequence both positions appear oversimplified and bear little resemblance to actually held views; yet so much lip-service is paid today to the ideology of historical understanding. Let me illustrate my criticism with examples taken from the otherwise interesting book by von Wright on explanation and understanding.[1]

Having distinguished in the history of ideas the Galilean (Platonic) tradition which identifies explanation with causal or mechanical explanation, and the Aristotelian tradition which emphasises primarily the value of teleological or final explanation, Wright characterises positivism as being in the Galilean tradition and as upholding the following three theses: (*i*) there is one scientific method common to all sciences, however diverse their subject matter, (*ii*) exact natural sciences, in particular mathematical physics, set a methodological ideal which measures the degree of development of all sciences, including the humanities, (*iii*) scientific explanations are causal, *i.e.* they consist in subsuming individual cases under hypothetically assumed general laws of nature.[2] Using 'positivism' in this way, which clearly amounts to methodological naturalism, Wright then classifies Hume, the philosophers of the enlightenment, the conventionalists Poincaré and Duhem, radical conventionalists Ajdukiewicz, Cornelius and

---

[1] Von Wright [1971].

[2] *Ibid.*, p. 4.

Dingler, logical empiricists and Popper as positivists and claims, among other things, that differences between them are negligible and criticisms 'somewhat patricidal' (*e.g.* Popper's criticism of logical positivism) or, perhaps, fratricidal (*e.g.* nineteenth-century positivists and conventionalists). By contrast to his opponents, his allies the 'anti-positivists', among whom he mentions Droysen, Dilthey, Simmel, Weber, Windelband, Rickert, Croce, Collingwood and Oxford linguistic philosophers, are given fuller attention and their diversity a more appreciative treatment. How adequate descriptively is Wright's account of 'positivism' and 'anti-positivism' in the nineteenth and twentieth centuries? A few examples should be sufficient to show its inadequacies:

Wright failed to mention that the conventionalism of Poincaré (and Duhem) originated from Kant's philosophy and that Poincaré believed in the synthetic *a priori* nature of some of our knowledge (*viz.* arithmetic) an anathema from the logical positivist viewpoint one of whose main tenets is the denial of the existence of synthetic *a priori* knowledge. He failed, moreover, to see that since neither Poincaré nor Duhem expressed any views concerning the relation between the natural sciences and the social science and humanities, neither thesis (*i*) nor (*ii*) in his characterisation of positivism may be relied on in classifying them as positivists. There remains thesis (*iii*). This was rejected explicitly, at least by Duhem, who argued that physical theories must not be regarded as explanations on pain of making physics dependent on metaphysics. When referring to radical conventionalism in general and to Ajdukiewicz's in particular as 'akin to positivism' he is obviously unaware of the fact that Ajdukiewicz explicitly propounded the 'anti-positivist' view of humanities in terms of their peculiar methods of 'understanding expressions' (hermeneutic) and of 'understanding the goal directed actions',[1] thus rejecting theses (*i*) and (*iii*). In reporting the views of the neo-Kantian Baden school, mainly of Windelband and Rickert, Wright emphasised one-sidedly the antipositivist element in their classification of scientific disciplines, namely the existence of idiographic[2] disciplines, while failing to mention the 'positivist' (naturalist) element, namely, Windelband's admission that there are idiographic and nomothetic disciplines among both natural and social sciences (*cf.* also Rickert's 'generalising' and 'individualising' sciences of nature *and* of culture). He disregarded the fact that Simmel owed and contributed perhaps as much to 'perspectivism', a biological and sociological account of the *a priori* elements in knowledge developed as a product of mutual influences between Comtian positivism, Spencerian evolutionism and neo-Kantianism, as did Nietzsche, Vaihinger and others. When discussing Dray's criticism of the nomological-deductive ('covering-law') schema of explanation and his account of the explanations in history in terms of reasons and rational actions, he failed to mention that Popper developed an account of explanation in history in terms of 'situational logic' and of 'rationality assumption', (*i.e.* in terms of the aims of the acting individual, his knowledge concerning the aims-means relation, the means in his possession, obstacles to be faced, *etc.*) along with an account of the nomological-deductive pattern of scientific explanations.

---

[1] Ajdukiewicz [1934] and [1949].

[2] Idiographic, *not* 'ideographic' (as Wright has it) which is etymologically different.

Wright has identified positivism with methodological naturalism (doctrine (*d*) on my list characterising strict positivism), as have many other philosophers of the social sciences and humanities. Such a concept of positivism has two defects. It is somewhat parochial and confusing. It is parochial since it is inapplicable to all those scientists and philosophers of science who have never expressed systematically any views on the methods and aims of the social sciences or humanities even though they may hold all the remaining views of positivism in the strict sense, *i.e.* doctrines (*a*) through (*c*), (*e*) and (*f*). Such a concept of 'positivism' is moreover confusing, since it does not allow for the fact that many critics of the remaining doctrines of positivism in the strict sense have been at the same time methodological naturalists, *i.e.* in agreement with positivism on just this one point and in disagreement otherwise, whereas at least some authors who accepted most doctrines of positivism, in the strict sense were at the same time *anti-naturalist* (methodological dualists), *i.e.* rejected (*d*). The terminological confusion helps thus to overlook the fact that in the history of philosophical doctrines there have been *positivist naturalism and antipositivist naturalism.*[1] Comte, the behaviourists (including Skinner today), Neurath, Carnap in his early (pre-1935) writings, for example, clearly represent the former; Popper and his school, for example, represent the latter.

The distinctions used here are not just more refined ways of discussing the positions of various philosophers. By emphasising the effect which different views of natural science had on the philosophers of social science and humanities, they suggest the following hypothesis concerning the origin of antinaturalism: 'Antinaturalism in nineteenth and twentieth centuries has been in general due to the acceptance by antinaturalist philosophers of either the positivist (phenomenalism, physicalism) or of mechanistic-determinist philosophy of natural science and to their refusal to extend the principles of those philosophical doctrines to the social sciences and humanities'. Without attempting to test this hypothesis systematically, I will mention a few cases which seem to satisfy it: Dilthey—as was pointed out by Richard von Mises[1]—was under the influence of Du Bois-Reymond who characterised the knowledge of nature in the following way:[2]

---

[1] I have made this distinction in my [1961] and in my [1964], pp. 149–60.

> ... Knowledge of nature—more precisely, the scientific explanation of the physical world by means of and in the sense of theoretical science—means reducing all changes in the physical world to the motions of atoms effected by time-independent central forces, or analyzing all natural occurrences into a mechanism of atoms.

It is not surprising that Dilthey, who accepted this mechanistic view of natural science, insisted on the impossibility of reducing cultural phenomena to facts of a mechanistic nature and, consequently, on the dualism of natural science and humanities. Similarly, when reading Winch's arguments against Mill's naturalism, one cannot fail to see that Winch accepted without questioning Mill's dated, nineteenth-century positivist account of natural science in general and of physics in particular. Winch insisted, for example, that—by contrast to the practice of natural science—explanations of human actions are not and cannot be mechanistic, that an adequate description of a human action cannot possibly be purely phenomenalistic, that for many individual human actions, or their products (*e.g.* of creative activities) there are no deterministic laws from which they could be predicted.[3] The fact that twentieth-century physical theories are neither mechanistic, nor deterministic nor phenomenalistic, seems to have escaped Winch's attention. Finally, one cannot help thinking that Wright's own anti-naturalism may be due to a large extent to his belief that natural science, physics in particular, was and still is in what Wright calls the Galilean tradition, which allows only mechanistic explanations. We may conclude that either Dilthey's, Winch's and Wright's anti-naturalism (methodological dualism) was mistaken, or—at least—that some of their reasons for rejecting naturalism may be seen today as wrong.

It appears, then, that while both Popper and Wright are anti-positivists, since they reject most of the doctrines (*a*) through (*f*), Popper holds the naturalist (monist) whereas Wright the antinaturalist (dualist) view of the social sciences.[4] It appears so, at least, if we judge from their explicit answers to the question whether or not natural and social sciences have the same aims and use the same methods. How do they characterise those aims and methods in detail, however?

The main thesis of Wright's book on explanation and understanding was formulated by him as follows:[5]

---

[1] von Mises [1939].

[2] *Ibid.*, p. 207.

[3] Winch [1958], pp. 73, 85, 87.

[4] Popper [1961], ch. 29 'The unity of method'.

[5] von Wright, *op. cit.*, p. 27.

> . . . It is a tenet of the present work that the practical syllogism provides the sciences of man with something long missing from their methodology: an explanation model in its own right which is a definite alternative to the subsumption-theoretic covering law model. Broadly speaking what the subsumption theoretical model is to causal explanation and explanation in the natural sciences, the practical syllogism is to teleological explanation and explanation in history and the social sciences.

Practical syllogism is, according to Wright, the schema (*PI*) of practical inference:

(*PI*) *A* intends to bring about *p*
A considers that he cannot bring about *p* unless he does *a*
Therefore *A* sets himself to do *a*.

The main characteristic features of practical inference so understood are, firstly, that the major (first) premise specifies the intention of a person at a certain moment to bring about a state of affairs; secondly, that the minor (second) premise indicates his belief as to what may constitute a necessary or sufficient condition for bringing about the intended state of affairs; thirdly that the 'validity' or correctness of the explanation (the explanandum is in the conclusion) does not depend on the correctness of the belief, *i.e.* on the truth of the sentence which claims that *a* is a necessary (sufficient or both) condition to bring about *p*; fourthly that the relation between *A*'s intentions and reasons, on the one hand, and his actions, on the other, may be regarded not necessarily as causal but as conceptual (*A* conceptualises *a* as a means to achieve *p* and thus also associates a positive value with *p*).

How does Popper characterise the methods of the social sciences and of humanities (*e.g.* of history)? Here is a brief account of explanations in history:[1]

> . . . most historical explanation makes tacit use, not so much of trivial sociological and psychological laws, but of what I have called . . . the logic of the situation; that is to say, besides, the initial conditions describing personal interests, aims, and other situational factors, such as the information available to the person concerned, it tacitly assumes, as a kind of first approximation, the trivial general law that sane persons as a rule act more or less rationally.

The fact that social scientists can make use of the assumption of rationality (that people act more or less rationally) is, according to Popper, very important for the methodology of the social sciences and of history:[2]

> . . . The last point [*i.e.* the rationality assumption] seems to me, indeed, to indicate a considerable difference between the natural and the social sciences . . . I refer to the possibility of adopting, in the social sciences, what may be called the method of logical or rational construction, or perhaps the 'zero method'.

---

[1] Popper [1966], vol. 2, p. 265.

[2] Popper [1961], p. 141.

> By this I mean the method of constructing a model on the assumption of complete rationality (and perhaps also on the assumption of the possession of complete information) on the part of the individuals concerned, and of estimating the deviation of the actual behaviour of the people from the model behaviour ... An example of this method is the comparison between actual behaviour and model behaviour to be expected on the basis of 'pure logic of choice', as described by the equations of economics ....

Popper's comparison of history with theoretical social sciences is based on the same principle as Rickert's distinction between individualising and generalising sciences, though Popper recommends 'the technological approach' as particularly useful in selecting problems and in imposing 'a discipline on our speculative inclinations'.[1] To be sure, Popper insists that not all human and social phenomena may be explained in terms of aims, motives, intentions, the individual's knowledge or prejudice, *etc.*, since many of those phenomena are in fact unintended consequences of the activities of numerous individuals making up a given society. I do not doubt, however, that Wright would not deny that both social scientists and historians have often to deal with and explain unintended consequences of both individual actions (due perhaps to misinformation) and of the actions of many individuals.

Are the views of positivists in the strict sense concerning explanations radically different from Wright's or Popper's?

In his 'Concise textbook of positivism' Richard Mises discusses critically various claims made by antinaturalists and clarifies the relation between 'positive' and normative science of nature and of society as follows:[2]

> ... One usually regards as the content of a positive science statements that assert something about *relations among phenomena*. Such a statement can be given the form: As soon as there is A, there is B. This may in suitable cases be replaced by the sentence: If B is desired, one has to see to it that A occurs ....

The following schema (*NI*) of normative inference is clearly implicit in the quoted passage:

(*NI*) *a* is desired (at time t by person *A* or group *G*) *p* is necessary to bring about *a*
Therefore *p* ought to be brought about (or in more explicit form: If *a* is desired, then *p* ought to be realised).

The main difference between (*NI*) and Wright's (*PI*) is in the minor premise which in (*PI*) refers to the actor's belief concerning the ends-means relation between *p* and *a*, whereas in (*NI*) it categorically affirms the ends-means relation between *p* and *a* presumably on the basis of some law borrowed from a positive science (natural or social, *e.g.* physics or economics). This difference is understandable since (*PI*) is claimed to be used in explaining actions, whereas (*NI*) is to provide instrumental norms.

---

[1] Popper, *op. cit.*, p. 58 ch. 20, 'The technological approach to sociology'.

[2] von Mises, *op. cit.*, p. 247.

John Kemeny, a philosopher of science whose views are close on many points to those of Carnap after 1935, *i.e.* to logical empiricism in its mature stage of development, characterised social sciences in a similar way. According to Kemeny the problem of characterising social sciences is intimately connected with the problem of how science is related to ethics, since the fundamental question a philosopher of science has to answer here is whether scientific problems in the social sciences can be divorced from questions of value. Having classified scientific (and mathematical) statements into analytic and synthetic (factual) and distinguished them from value statements, Kemeny re-affirmed the principle that no value statement can be deduced from scientific (factual or analytic) statements alone. Then he described the relation between value statements and science using an example of inference whose structure is basically the same as that of (*NI*):[1]

> . . . let us suppose that we have decided that eradicating multiple sclerosis from the US is a highly desirable goal. We . . . (then) . . . consult the biological scientists and physicians who are best qualified to give us advice . . . If we try to schematize the type of argument used, it may look somewhat as follows: 'It is good that multiple sclerosis should be eradicated from the US. The following is the list of actions that are necessary to eradicate multiple sclerosis from the US. Therefore these actions should be undertaken' . . ..

In general the relation between value statements and science is the same: science can tell us what decisions are open to us; moreover ' . . . it can tell us what the future will be like, depending on the way we make our decisions. Then we may think of value statements as choosing between possible future states . . . '.[2] Now, turning to the social sciences Kemeny claims:[3]

> . . . It cannot be denied that it is in principle feasible and in practice important to find scientific laws which will enable us to correlate human means to human ends. This type of activity is appropriately described as social science . . ..

Is there perhaps a fundamental difference between the naturalist and antinaturalist conceptions of humanistic explanations in terms of whether the humanistic terms like 'intends', 'prefers', believes', 'aims at maximizing', *etc.* are understood in the instrumentalist or realist fashion? This suggestion may be found in the writings, concerned with the peculiarities of the social sciences and humanities, of some Polish (marxist) philosophers,

[1] Kemeny [1959], p. 235. [2] Kemeny, *op. cit.*, p. 238. [3] Kemeny, *op. cit.*, p. 246.

the so-called 'Poznan group'.[1] They have accepted my distinction between positivist and antipositivist naturalism and claim that marxism, as understood by them, takes the naturalist but anti-positivist position. Their account of explanations peculiar to social science and humanities is, in brief, as follows:

Humanistic explanations are answers to questions such as: 'Why has X performed action C?', 'Why does the outcome (product) of X's action have the property C?', *etc.* Such explanations are in terms of a set of alternative actions $C_1, \ldots C_n$ contemplated by the agent X, in terms of his beliefs as to the outcomes $S_1, \ldots S_k$ of those actions, in terms of X's ordering the outcomes with a preference relation R and in terms of the assumption of rationality (implicit rather than explicit in most cases). As regards their logical form, humanistic explanations fall under the general nomological-deductive pattern; their peculiarity is to be seen in the structure determined by the extralogical terms. The terms in humanistic explanations are relativised to the agent's knowledge and value system; they are, in this sense, subjective; the terms denoting alternative actions are intentional. All these terms are theoretical (non-observational) as are the premises in which they occur and which, therefore, have to be tested indirectly; moreover, the premises concerning the agent's knowledge and preferences are idealised (and, therefore, theoretical also on this account). Since among the premises of humanistic explanations there are theoretical sentences, the three views concerning the cognitive status of theories, *viz.* descriptivism, instrumentalism and realism, are applicable to them. On the descriptivist (phenomenalist or positivist) view the peculiarly humanistic terms are eliminable in favour of observables (*cf.* behaviourism). According to the realist view, the premises of humanistic explanations concerning the agent's beliefs and preferences, refer (hypothetically and approximately) to the agent's actual states of consciousness (knowledge, valuations, *etc.*). On the realist view of the rationality assumption, the latter describes (hypothetically) an actual regularity, *viz.* definite relation between the agent's states of consciousness and the actions undertaken by him. By contrast, instrumentalists deny the realist's claims, regard sentences with subjective terms as mere instruments for making predictions in conjunction with suitable correspondence rules and allow subjective terms to be related via correspondence rules (which need not be explicit definitions) not only with human actions but also with the behaviour of lower animals and with certain observable states of automata. (Note, that 'instrumentalism' is used here not in Popper's sense but in a sense closer to 'moderate instrumentalism' as characterised in the previous section).

[1] Kmita and Nowak [1968]; Zamiara [1974].

'From the realist viewpoint agents are (more or less) conscious of the elements of the decision-situation in which they make decisions, they foresee some of the outcomes of their decisions and choose those outcomes which they, in given circumstances, most prefer'.[1] *The peculiarity of humanistic* explanations consists in the use of subjective terms and sentences (concerning beliefs, preferences, *etc.*) in a realist fashion.

It is not clear but rather doubtful whether the 'realist view of humanistic explanations' of the Poznan group would be acceptable to Wright, since it seems close to the 'causalist' position rejected by Wright, who regards explanations in social science as 'quasi-causal' and who explicitly claims that 'intentionality is not anything "behind" or "outside" the behaviour. It is not a mental act or characteristic experience accompanying it.'[2] Unless, however, he provides the humanistic terms with some realistic (though not necessarily mentalistic) interpretation, no positivist would object to re-describing behaviour in intentional terms; even descriptivists permit theoretical re-description as an equivalent way of talking about observables; non-equivalent theoretical (*e.g.* intentional) re-description is certainly within a moderate instrumentalist's rights (*cf.* Duhem and Poincaré). All this indicates alternative explications of 'realism' with respect to humanistic statements are required, preferably in connection with specific psychological theories involving 'intends', 'believes', *etc.* In any case, it seems unlikely that the distinction between realism and instrumentalism with regard to humanistic terms will coincide with the distinction between naturalism and antinaturalism.

My survey of the five accounts of the aims and methods of the social sciences and humanities by Wright, Popper, Mises, Kemeny and the 'Poznan group' shows, I think, that—insofar as they are representative of the writings of many antipositivist antinaturalists (Wright), antipositivist naturalists (Popper, Kemeny, marxists), positivist naturalists (Mises)—there is in fact more agreement on the subject between the various parties concerned than appears from their polemical comments occasionally hurled at one another. The widespread belief that it is otherwise seems to me to rest, to a large extent, on misunderstandings, in particular on the following three: (1) Contrary to the claims of antinaturalists (Winch, Wright, for example), none of the naturalists in my survey claims that descriptions in the social sciences (and humanities) are or ought to be purely phenomenalistic, *i.e.* in behaviouristic terms, that explanations of human actions are or ought to be mechanistic, without reference to human aims, beliefs, *etc.* Just the opposite is the case: they all insist that social sciences and humanities are concerned primarily with studying men as rational beings, *i.e.*—to use Antony Flew's words[1]—'. . . as adopters and abandoners of beliefs, pursuers and frustraters of purposes, makers of plans and so on . . . '; they all agree that typical explanations of individual actions in history (and humanities in general) are in terms of aims, intentions, beliefs, available means, existing obstacles, institutional set-ups, *etc.* They also agree that one of the methods available to social scientists and used by them is the construction of models of decision-making based on various assumptions, including very often the assumption of (perfect) rationality and of perfect knowledge—finally—that a practical, technological approach concerned with studying the relation between human ends and means is the most fruitful heuristically and methodologically. The claims of antinaturalists in this case result from mistaken identification of naturalism with behaviourism or physicalism and from their failure or refusal to acknowledge the fact that changes have occurred in the philosophy of those they attack (*e.g.* Carnap's views on theories before 1935 and after, especially in his article on 'The methodological character of theoretical concepts', 1956) just as in their own philosophy (*e.g.* Wittgenstein's *Tractatus* and *Philosophical Investigations*). (2) When naturalists affirm and antinaturalists deny that the aims and methods of the social sciences and humanities are basically the same as the aims and methods of the natural sciences, the former understand 'the aims and methods' in very general terms (*e.g.* 'to predict and explain, using empirical hypotheses') whereas the latter use the same phrase to stand for very specific features of 'the aims and methods' (*e.g.* 'to explain individual actions in terms of intentions, beliefs, *etc.*'). In effect, the disagreement is more apparent than real. (3) More specifically, though Wright insists that the validity of explanations falling under his schema (*PI*) does not depend on the truth of the generalisation concerning the ends-means relation mentioned in the minor premise, whereas the validity or adequacy of a nomological-deductive type of explanation does depend on the truth of the laws involved, one should point out the fact that the validity of the explanations of type (*PI*) does depend on the truth of *some* generalisation, namely on the truth of the following generalisation (*G*) which is not a logical tautology and may run somewhat as follows:

---

[1] Zamiara, *op. cit.*, p. 148.

[2] von Wright, *op. cit.*, p. 115.

(*G*) If x intends to bring about Q, and believes that he cannot bring about Q unless he does R, and that he can bring about R, then x sets himself to do R.

From an instantiation of (*G*) by suitable substitutions and the conjunction of the major and minor premise of (*PI*), the conclusion of the inference (*PI*), *i.e.* the explanandum, is logically derivable in virtue of the rule of detachment. If so, then explanations falling under schema (*PI*), regarded by Wright as peculiar to social sciences and humanities, are in fact a proper subset of the set of nomological-deductive explanations. To insist on their importance in the social sciences and humanities amounts to a demand that nomological-deductive explanations in those areas should be studied in greater detail than had been done by Hempel and others. No naturalist need object to such a programme of methodological research. If (*G*), though not a tautology, is a meaning-postulate for 'intends', 'believes', then it still may occur as an explanatory premise, provided that there are other postulates relating the two terms with other observables.

---

[1] Flew [1972].

Finally, even if the last mentioned condition is not satisfied, (*G*) may still be regarded—at least from the moderate-instrumentalist (*e.g.* conventionalist) viewpoint—as a theoretical statement re-describing observable behaviour in intentional terms. (*G*) would then be comparable to physical theories which re-conceptualize observables in a more abstract and idealised language of scientific metaphors, conventions and fictions rather than provide causal explanations of the phenomena. Thus the claim that the link between action and intention as described in a humanistic explanation is of a ceneeptual rather than causal nature, turns out to be very close to instrumentalism (conventionalism) and thus to naturalism in its instrumentalist version. This is not surprising since both originate from the neo-Kantian philosophy.

4 POSITIVISM AND POSITIVE ECONOMICS

In the present antipositivist climate it would be surprising indeed if the methodological conception of economics as a positive science (positive economics for short) were not to become the target for antipositivist attacks. Its very name, of Comtian origin, is sufficient to provoke suspicions. In due course, after the antipositivist tendencies had gathered momentum everywhere, came its turn to stand accused of positivist vices.

But what is positivist about positive economics, apart—that is—from its name? This is what we want to clarify, also to examine what—if anything—is wrong with economics being positive.

There are many accounts of the aims and methods of economics as a positive science. Any textbook in the neo-classical tradition will provide one. John Neville Keynes's discussion deserves special attention perhaps, since he seems to have popularised the Comtian term among economists. Among more recent publications, T. W. Hutchison's book on the basic postulates of economic theory or R. G. Lipsey's textbook are often mentioned as paradigmatic. However, none of these has been as much discussed as Milton Friedman's essay on 'The methodology of positive economics'.[1] It provoked a long and heated controversy in the nineteen sixties and continues to attract criticisms now. In the nineteen sixties, the methodological controversy centred around the question of the admissibility of 'unrealistic assumptions' in economic theory; the basic principles of the methodology of positive economics were not questioned. In more recent criticisms, it is those principles themselves (as well as the advocacy of unrealistic assumptions) that have been under attack.

I shall use here Friedman's account of the methodology of positive economics to examine the relevance of antipositivist criticisms. In order to isolate Friedman's individual views from the principles essential to positive economics, I shall rely on the differences between the controversy in the nineteen sixties and recent antipositivist criticisms.

At the outset of his essay Friedman reminded us that John Neville Keynes had distinguished 'positive science' as '. . . a body of systematic knowledge of what is . . . ' from 'normative science', *i.e.* '. . . a body of systematized knowledge discussing criteria of what ought to be . . . ' and from 'art', *i.e.* '. . . a system of rules for the attainment of a given end . . . '. Having adopted these distinctions, Friedman formulated certain principles and claims concerning economics (as a science) which I shall summarise in the following seven points.[2]

(1) *Positive economics*, by contrast to normative science and art, deals with 'what is' not with 'what ought to be'; in this sense it is, in principle, independent of any ethical position or normative judgment. Its task is to provide a system of generalisations that can be used to make correct predictions; its performance is to be judged by the precision, scope and conformity with experience of the predictions it yields.

Positive economics can thus be objective in the same sense as any of the physical sciences; the fact that it deals with interrelations of human beings and that the investigator himself is one, raises special difficulties in achieving objectivity while at the same time it provides the social scientist with data not available to the natural scientist; neither the one, however, nor the other makes it fundamentally different from natural science.

(2) *Economic policy*, if rational, rests on predictions about the consequences of doing one thing rather than another; those predictions, in turn, are based implicitly or explicitly on positive economics; there is, of course, no one-one relation between policy conclusions and the conclusions of positive economics, since policies may be concerned with achieving different aims (values).

[1] Friedman [1953].

[2] Friedman, *op. cit.*, pp. 509, 512, 513, 516–17.

(3) *Economic theory* may be seen as consisting of two elements: a language or system of concepts for organising empirical data, and a body of substantive (empirical) hypotheses. Viewed as a body of hypotheses, theory is judged by its predictive power for the class of phenomena which it is intended to 'explain'. A hypothesis can never be verified. It is rejected if frequently contradicted by evidence, accepted otherwise. Greater confidence is attached to it, if it has survived many opportunities for falsification. The validity of a hypothesis in this sense is not sufficient for choosing among alternative hypotheses, many of which may be consistent with available data; other criteria are then used, *e.g.* simplicity, fruitfulness, *etc.*

(4) *The testing of predictions* in economics, unlike in most natural sciences, is not generally done. The difficulty, due to the impossibility of controlled experiments, of getting new *evidence* necessary to test the observational consequences ('implications') of theories, has prompted economists to distinguish what they call the 'assumptions' of a theory from its 'implications' and to claim that '. . . the conformity of these "assumptions" to "reality" is a test of the validity of the hypothesis *different* from or *additional* to the test by implications . . . '.

(5) *The development of the theories of imperfect competition* was motivated by and their acceptance based on the belief that the assumptions of 'perfect competition' and of 'perfect monopoly' (said to underlie neo-classical economic theory) are descriptively false. Later, similar claims of the falsity of assumptions were made in the controversy over marginal analysis.

(6) Insofar as it is possible to isolate the assumptions of a theory and directly assess their accuracy, 'the relation between the significance of a theory and the "realism" of its "assumptions" is almost the opposite of that suggested by the critics of unrealistic assumptions', namely truly *important theories have assumptions* which are *'wildly inaccurate descriptive representations of reality'*. A theory, namely, is valued for 'explaining' much by little. Therefore, since a theory has to be abstract, its significance varies inversely with the 'realism' of its assumptions.

(7) The 'validity' of a theory must be judged not by assessing the 'realism of its assumptions' but by the conformity of its consequences with the facts which the theory predicts or explains.

The preoccupation with the 'realism of assumptions' rather than with empirical testing produces '. . . a misdirection of much intellectual effort . . . and impedes the attainment of concensus on tentative hypotheses in positive economics'.

Significantly, in the dispute over Friedman's essay in the nineteen sixties none of the listed views, except (6) and (7), were at issue, *i.e.* either criticised or defended. Evidently the dispute was on the question of the admissibility and usefulness (or otherwise) of 'unrealistic assumptions' in economic theory. Claims (6) and (7) which formulated Friedman's instrumentalist view on this point and that of other economists sharing his position were criticised either from the position of descriptivism (*e.g.* Samuelson), or realism (*e.g.* H. Simon, J. Melitz); they were defended by some instrumentalists (*e.g.* F. Machlup), critically clarified by other instrumentalists (*e.g.* G. Massey) or from a neutral philosophical position (E. Nagel). Claims (1) through (3) were implicity assumed by all disputants, while the falsificationist content of (3) received support from some of them (*e.g.* K. Klappholz and J. Agassi).[1] I assume, therefore, that only theses (1) through (3) are essential for the methodology of positive economics (theses (4) and (5) are of diagnostic nature) and that they are to be understood as general enough to allow for different epistemological views concerning the status of theoretical terms and statements and for different views concerning the best strategy in economic research, *e.g.* for strategies based either on the single-theory model or on the multi-theoretic model.

Since I have just classified Friedman's view of economic theories as instrumentalist, some qualifying comments are in order. Unlike Machlup, Friedman never said explicitly that theoretical sentences are neither true nor false, and in (1) he explicitly denied that there was no methodological difference between pure (positive) and applied economics (normative science and art). It would appear, therefore, that Popper's definition of instrumentalism is inapplicable to his position. In fact, Friedman's instrumentalism consists in (*a*) his praise of 'unrealistic assumptions' in theorising, (*b*) his emphasis on the predictive power as the main or even the only criterion for the 'validity' of a theory, (*c*) his claim that psychological investigations concerning the question whether or not agents in fact act rationally in the economist's sense (or use marginal analysis) are irrelevant as long as theories based on the assumption of profit maximisation yield accurate predictions. The last claim amounts to an instrumentalist use of 'subjective' terms ('aims', 'intends', 'believes', *etc.*) as 'mere instruments of prediction'. The first, (*a*), seems to be a reminiscence of the 'fictionalism' of neo-Kantians (*e.g.* F. Lange, H. Vaihinger) and of conventionalism (H. Poincaré, P. Duhem). It was the neo-Kantians who claimed that all knowledge is conceptualised and in this sense subjective, that abstract concepts necessarily 'falsify' and 'distort' reality which—in itself—is unknowable, that our language used both in everyday life and in science is full of fictions, metaphors and anthropomorphisms. To quote Nietzsche in his neo-Kantian mood: 'If (man) does not wish to be satisfied with truth in the form of a tautology—that is, with empty shells—then he will forever buy illusions for truths'.[1] It was the conventionalist Poincaré who claimed that physical theories are couched in metaphorical language (and a metaphor, taken literally, is a false sentence or one incompatible with accepted knowledge) and the conventionalist Duhem who contrasted physical laws as highly idealised and symbolic representations with commonsense laws which are true in the straightforward sense. At the same time, these philosophers regarded intellectual fictions and conventions as useful, indispensable in our struggle for self-preservation and even as the only source of certainty. Through the French neo-Kantian Charles Renouvier the neo-Kantian doctrines influenced William James (who rejected the classical concept of truth in favour of the pragmatist) and other American pragmatists. The F-twist, as Samuelson nicknamed it,[2] is an echo of those nineteenth-century half-truths. On the other hand, Friedman's insistence on tenacity in empirical research, *i.e.* his claim that one ought to concentrate on the established theory (of perfect competition) as long as its rivals are not better at providing predictions is alien not only to nineteenth-century conventionalism but also to the ancient instrumentalist tradition in science (astronomy) which advocated multiplicity of rival observationally equivalent theories. In this respect, the use made by Friedman of principles (6) and (7) is rather akin to the role played in Newton's philosophy by Rule IV, which advocates the mono-theoretic strategy in science.

---

[1] P. Samuelson, H. Simon, F. Machlup, G. Massey, E. Nagel contributions to the discussion on the Problems of Methodology, *American Economic Review*, 1963. Klappholz and Agassi [1964].

Let us turn now to more recent criticisms of the methodology of positive economics. Characteristically, they are antipositivist and concerned not only—and not primarily—with the issue of 'unrealistic assumptions' but with questioning the basic claims (1) through (3). For example, A. Coddington[3] in an attempt '. . . to inquire into the present state of (positivist) tradition' in economics and motivated by the desire to clarify '*how* we are to be empiricists', referred to Friedman and Lipsey as leading representatives of the methodology of positive economics—that 'unduly rigid and stultified' form of empiricism.

Aware of some of the intricacies of the history of positivism, Coddington comments: 'The philosophers having positions closest to those of Friedman and Lipsey are probably William James and Karl Popper. Neither of these are positivists'. We are then given a characterisation of the methodology of positive economics in terms of the following four statements (*i*)–(*iv*). (*i*) 'The characteristic of a theory which alone is relevant in appraising it as a contribution to economic knowledge is its predictive performance', (*ii*) 'The primary yardstick for the acceptance or rejection of a theory is the compatibility of its predictions with something which is variously referred to as "facts", "experience" or "observation" ', (*iii*) 'There is no fundamental methodological difference between social and natural science', (*iv*) 'Theory should be value-free'.

---

[1] F. Nietzsche 'On truth and lie in an extra-moral sense', trans. in Kaufmann [1968], p. 45.

[2] Samuelson, *loc. cit.*, p. 232.

[3] Coddington [1972].

In his criticisms of the methodology of positive economics so characterised, Coddington makes use of some of the antipositivist arguments that have been advanced in contemporary philosophy, *e.g.* by Winch, Kuhn, *etc.* His main critical points may be summarised as follows:

In virtue of (*i*), the essential role of theories as explanations is ignored. The identification of 'explanation' with prediction and the 'as if' instrumentalist interpretation of theories (as analytic statements) leave, in fact, no room for explanations 'in the customary sense' (*i.e.* as 'causal narratives') which requires a realist view of theories.

No analysis is given of what should count as evidence and how it is related to facts (states of affairs), experience (observational reports), observational statements. It is not clear why the weighing of evidence should be impersonal. Can the dichotomy between facts and theories be maintained? If not, as Friedman somewhat inconsistently seems to admit, then—since there are alternative theories—can tests be objective? Why should theories always yield to facts? Is the conception of testing (theories by observations) useful in the methodology of empirical research? Are tests always consequential (*i.e.* have unambiguous implications for the theory)? Serious arguments have been advanced in contemporary philosophy of science in support of the view that the answers to these questions should be negative.

The claim of the unity of method (*iii*) ignores the fact that, unlike in natural science, in social science one can either adopt a 'first-person (participant's) viewpoint' or a 'third-person's (observer's) viewpoint'. In fact, methodological unity is achieved only by restricting the allowable empirical basis in microeconomics to the third person's viewpoint, which—as has been argued (Winch)—is necessarily impoverished and misleading.

The principle of value-neutrality (*iv*) ('economic theory should be ethically neutral') is, in fact, violated by the view that language of economic theory (neo-classical economics) is something 'given'. The positivist 'unqualified emphasis on the testing of theories *via* their predictions suggests that the conceptual framework is regarded as perfectly satisfactory' 'the demand for a justification of substantive assertions requires that the framework in which they are expressed and the criterion of justification are themselves accepted without justification'.[1] But the language of economic theory is both 'a spring-board' and 'a straight-jacket' and theoretic development may consist in attempts 'to fight one's way out of the straight-jacket'.

In attempting to evaluate the reported critical comments one faces various difficulties. It is not quite clear what bearing these critical comments have on the subject under dispute, *i.e.* the methodology of positive economics. They express disapproval of that methodology; that much is obvious. Some of them express preference for a programme of economic research different from Friedman's: Friedman likes theories which give accurate predictions and does not care much about anything else while others want theories which provide explanations (perhaps in terms of 'causal narratives').

One wonders, however, whether it would be at all proper to consider seriously the question: Have those critical comments established (or given support to) Coddington's main thesis or rather objection against the methodology of positive economics? But what is the main objection? That this methodology is 'unduly rigid and stultified' as a form of empiricism? Surely, whether it is or isn't unduly rigid depends on how free in research one likes to be, and we saw a while ago, that Friedman's likes are Coddington's dislikes. It appears that there would be no clash, except in preferences, were it not for the fact that Friedman wants everybody to share his likes and Coddington wants also everybody to share his (Coddington's) likes. Well, we can't satisfy them both. Hence the dispute. Are their programmes not both unduly rigid? Is there not enough room for both types of research?

Perhaps the criticisms were based on the implicit assumption that one of the aims of science is to explain (and that the aim of the social sciences is to explain in terms of aims, beliefs, *etc.*) and that this aim cannot be achieved following Friedman's instrumentalist methodology. But the aims of science are either somebody's aims—and then we return to the point that Friedman's aims may be different from the aims of other economists—or else they are a theoretical (philosophical) construct used somehow to account for certain features of science; in this latter case, the aims of science as a theoretical (philosophical) construct will vary with one's theory (philosophy) of science. Although many antipositivists today echo Duhem (and other conventionalists) in claiming that neither experiments nor any rational arguments can conclusively decide between rival scientific theories, they have surprisingly great confidence in the logical power of their arguments to decide between philosophical doctrines.

---

[1] Coddington, *op. cit.* p. 15.

Though there are few arguments in Coddington's criticisms of the methodology of positive economics, let us briefly look at those criticisms.

His criticisms are based, I think, on an inaccurate and misleading reconstruction of the principles of that methodology. It is inadequate and misleading since it fails to separate those methodological principles which are peculiar to Friedman's individual position from those essential for positive economics (shared by other disputants in the controversy over 'unrealistic assumptions' in the nineteen sixties). Friedman's and Lipsey's views were mentioned as representative for the methodology under discussion. If it is true—as Coddington claims—that in epistemology and methodology Friedman's views are close to James's and Lipsey's to Popper's, then since philosophically James and Popper are worlds apart—in particular, the former is an instrumentalist and the latter an anti-instrumentalist—provision should have been made for such a divergence in reconstructing the principles of the methodology. But no allowance has been made for a realist version along with the instrumentalist one. As a result, some of the criticisms directed against the methodology of positive economics in general, are irrelevant to its realist version. This applies to the first and third of Coddington's criticisms as reported here.

The realist use of subjective terms ('intends', 'aims at maximising', 'chooses', 'believes', *etc.*) which can easily be exemplified by references to Lipsey's textbook (*e.g.* '. . . We assume that anything we observe in the world and any relationship we assume in our theories can be traced back to decisions taken by individuals or groups')[1] is sufficient to satisfy one's desire for intentional (humanistic, 'first-person') microeconomic explanations. A realist interpretation of all assumptions of economic theory allows for explanations in the customary scientific sense of nomological-deductive explanations. The claim repeated by Coddington after Winch that 'external' or 'third-person' accounts of human behaviour somehow 'distort' human and social facts and that, therefore, only explanations of actions in terms of agents' own concepts are admissible, has not been supported by any convincing arguments and in any case would not be consistent with Coddington's insistence on the essentially causal nature of explanations in the social sciences, if the relation between actions and intentions is regarded as conceptual rather than causal.

Coddington's third criticism—one of the most often repeated in disputes over positive economics—is based on a suspicion, on what the principles 'suggest' rather than on what those principles explicitly state or recommend.

[1] Lipsey [1974], pp. 67, 236–7.

Of course, when one talks about test-statements and about testing theories against such statements, one assumes a language for those theories and test-statements as more or less given. Similarly, having decided which problems are of interest and importance (to oneself and to people sharing one's concerns) one will use a language which one thinks is most suitable for those purposes. Since most researchers, like most people in general, are traditionalists rather than revolutionaries, tradition plays a considerable role in the 'choice' of the language. The concept of 'objectivity', 'objective testing', *etc.* are, it is true, 'immanent concepts' (in Quine's sense),[2] *i.e.* definable for a given language, just as the concept of rational behaviour is relative to given ends or aims. But there is nothing in the principles of the methodology of positive economics to prevent one from criticising the choice of problems and the choice of language (with its underlying assumptions or postulates) just as there is nothing in the concept of rational choice of means to prevent one from criticising the choice of ends (from another viewpoint, *e.g.* as means to further ends). The fact that the principles of the methodology do not contain any encouragement to such criticism points out only their immanent nature. However, there is nothing unduly rigid about this since they can be easily supplemented by further recommendations for what one regards as fruitful research policy from the viewpoint of one's interests. Friedman's emphasis on the fact that there is no one-one relation between policy conclusions and the conclusions of positive economics in view of the role of norms or ultimate values (which are not derivable from positive economics), clearly acknowledges the function of values as guiding the choice of problems and of suitable linguistic frameworks. Moreover, since those principles permit both a realist and an instrumentalist version, it is worth mentioning that the instrumentalist and conventionalist tradition have always been associated with the policy for 'proliferation' of theories and languages. Conventionalists, both in the nineteenth century (Poincaré, in particular) and among logical empiricists (Carnap, Ajdukiewicz), were the first to consider the problems of the choice of a language and radical conventionalists (LeRoy and Ajdukiewicz) were the first to consider languages which are not intertranslatable.

Coddington's criticisms of principle (*ii*) on his list, concern the descriptive inadequacy of the methodology of positive economics (as an account of how empirical research in economics is actually done) rather than its normative consequences. As such, they are not really relevant to his conclusion that the methodology is too restrictive as a form of normative empiricist philosophy. He fails to note, however, that various refinements in the account of the relation between theories and evidence had been discussed extensively by various philosophers of science in the empiricist tradition, especially in the disputes over the conventionalist variety of empiricism. One must not expect from Friedman's formulation of the principles of positive economics phrased in very general terms and for a specific polemical purpose to refer to all those detailed refinements; though in fairness to Friedman one ought to note his emphasis on the difficulties of interpreting and assessing evidence in economics.

---

[2] Quine [1970], pp. 19–20.

University of Sussex

REFERENCES

Adjukiewicz, K. [1934]: *Logiczne podstawy nauczania* (Logical Foundations of Education) (Warszawa).

Adjukiewicz, K. [1949]: 'The Scientific World-perspective' in H. Feigl and W. Sellars (*eds.*) *Readings in Philosophical Analysis* (New York).

Coddington, A. [1972]: 'Positive Economics', *Canadian Journal of Economics*, 5, p. 1.

Flew, A. [1972]: 'Anthropology and Rationality', *Question* (January), p. 91.

Friedman, M. [1953]: 'The methodology of positive economics' in *Essays in Positive Economics*, and reprinted in M. Brodbeck (*ed.*) *Readings in the Philosophy of the Social Sciences* (New York, 1968), pp. 508–28.

Giedymin, J. [1961]: *Naturalizm i antynaturalizm w pojmowaniu nauk spolecznych.*

Giedymin, J. [1964]: *Problemy, zalozenia, rozstrzygniecia*. Poznan.

Kaufmann, W. [1968]: *Nietzsche*, translated and edited by W. Kaufmann.

Kemeny, J. [1959]: *A Philosopher Looks at Science.*

Klappholz, K. and Agassi, J. [1964]: 'Methodological Prescriptions in Economics', *Economica*, **26.**

Kmita, J. and Nowak, L. [1968]: *Studia nad teoretycznynis podstawami humanistyki* (Studies in the Foundations of Humanities).

Lipsey, R. [1974]: *An Introduction to Positive Economics*, 3rd edition.

von Mises, R. [1939]: *Kleines Lehrbuch des Positivismus*, translated as *Positivism. A Study in Human Understanding* (New York, 1951).

Popper, K. R. [1961]: *The Poverty of Historicism.*

Popper, K. R. [1963]: 'Three Views Concerning Human Knowledge' in *Conjectures and Refutations.*

Popper, K. R. [1966]: *The Open Society and its Enemies.*

Quine, W. [1970]: *Philosophy of Logic.*

Winch, P. [1958]: *The Idea of a Social Science.*

von Wright, G. H. [1971]: *Explanation and Understanding.*

Zamiara, K. [1974]: *Metodologiczne znaczenie sporu o status poznawczy teorii* (The methodological significance of the controversy over the epistemological status of theories).

Brit. J. Phil. Sci. 26 (1975), 303–318 *Printed in Great Britain*

Theory Choice in a Two-level Science*

by GEOFFREY SAMPSON

A chief problem of the philosophy of science concerns the proper criterion or criteria for choosing between alternative, incompatible theories about given subject-matter. Thus, Popper [1934] argues that, among theories not refuted by observation, one should choose the most highly refutable or strongest; he uses the term 'simplicity' interchangeably with 'strength'. Goodman [1961] points out that simplicity and strength are distinct properties of theories, and he argues that both should independently be maximised (*cf.* Quine [1963], Rudner [1966], §§2.7, 2.9). Kuhn [1962] and Feyerabend [1970] treat the belief in rational criteria for theory choice as a delusion, claiming that such choices can be decided only by factors such as the relative charisma of the proponents of different theories. Finally, Lakatos [1970], while conceding that the 'goodness' of individual theories will not normally be commensurable, suggests that there do exist rational criteria for comparing and evaluating rival 'research programmes' or evolving series of theories: we should adopt 'progressive' and reject 'degenerating' research programmes.

These divergent views were worked out against a shared background of exposure to substantive scientific theorising; the writers mentioned, and other modern philosophers of science, draw heavily on physics, and to a lesser extent on chemistry and biology, in order to illustrate and support their theses. There also exists a body of writing on the philosophy of the social sciences, particularly economics, anthropology, and sociology; however, there is a difficulty in using these subjects as a basis for arguments about the nature of science in general, in view of the fact that even practising social scientists themselves are often in genuine doubt about whether their subjects rank as 'sciences' in a meaningful sense. It has been argued that the standard concept of 'social science' is inherently contradictory (Popper [1957], Winch [1958]).

The purpose of this paper is to show that modern linguistics (which, I shall suggest, is a science in the full sense of the word) involves a type of problem about theory choice which has not previously been discussed in

---

* In preparing this article for publication, I have profited greatly from criticisms of an earlier version by the Editors of this *Journal* and by participants at an L.S.E. seminar. Responsibility for its defects is mine alone.

Received 4 *February* 1975

the philosophical literature. The type of problem to be described is not intrinsically restricted to linguistics; I would expect that cases occur in other sciences, so that the discussion should be relevant for philosophers with no special interest in linguistics. But it seems to follow from a peculiarity of the methodological structure of linguistics that the problems to be discussed are particularly obtrusive in that science.

I begin with a brief sketch of the structure of linguistics as a science; readers are referred to my [1975] for a fuller account.

Linguistics includes two 'levels' of theories. At the lower level, it includes theories about individual languages. (Within linguistics, such theories are commonly called 'grammars'; I shall also call them 'first-order linguistic theories'.) A first-order linguistic theory, or grammar, consists of a definition of a set of word-sequences (normally an infinitely large set) preceded by a statement that the sequences in the set are all and only the sentences of some natural language L. Thus, an observation falsifying a grammar G_L of a language L will be of the form 'On occasion t, L-speaker p uttered word-sequence S', where S is not in the set defined by G_L. (For instance, the sequence 'Is the if was could' might be outside the set of sentences permitted by a well-supported grammar of English.) At the higher level, linguistics includes rival 'general theories of language', or 'second-order linguistic theories'. A second-order linguistic theory defines a class of sets of word-sequences (again, normally an infinitely large class) and claims that all and only the sets in this class can occur as natural human languages. Thus, an observation falsifying a second-order linguistic theory T will be of the form 'The mother tongue of linguistic community c consists of the set L of word-sequences', where L is not in the set defined by T. (For instance, for any vocabulary V of words, the set $\{x|x$ is a sequence of words drawn from V and the length of x is prime$\}$—let us call this the 'prime-number language for V'—might be outside the class of sets of sequences permitted by a well-supported second-order linguistic theory.) In practice, a general theory of language works by specifying a class of grammars and stating that the languages which humans could use are just the languages generated by grammars in the specified class.

As in other sciences, we must understand the empirical predictions made by linguistic theories to be modified by implicit *ceteris paribus* clauses. In the case of second-order linguistic theories such clauses will perhaps only rarely be relevant, but in the case of grammars of individual languages they will certainly need to be invoked quite frequently. Thus, one would not take a grammar of English which forbade the sequence 'Is the if was could' to be refuted by an observation of an English-speaker uttering that sequence, if, for instance, the speaker was being bribed to utter it or threatened with unpleasant consequences if he did not utter it (or if the speaker were a linguist wishing to exemplify the notion 'ill-formed sequence').

Partly because of failure to understand the role of *ceteris paribus* clauses in scientific theories, Noam Chomsky, the creator of the modern linguistic research programme,[1] has argued that grammars must be interpreted as making predictions not about speakers' utterances but about their intuitive knowledge of their language. There is a sense of 'intuition' in which a scientific theory can be based on intuition and yet be empirical. I might, for instance, base some physiological theory on, *inter alia*, the datum that people perspire more when physically active than when at rest; but if I were challenged as to what evidence this latter statement is based on, I would have to admit that I cannot cite evidence for it—for me, at least, it is an intuitive belief subconsciously distilled from many past observations to which I paid little attention at the time and which I have since forgotten. However, an intuitive belief in this sense is just as much open to empirical testing as is a proposition which has been consciously formulated as a tentative scientific hypothesis. Chomsky holds grammars to be based on intuition in a more objectionable sense. For him, a native speaker's beliefs about the well- or ill-formedness of sequences of words in his language are in principle the ultimate criterion for the acceptability of a grammar of that language. A speaker's overt linguistic behaviour will mirror his intuitive knowledge of his language only imperfectly, and, where intuitions and behaviour conflict, Chomsky treats intuitions as decisive for accepting or rejecting a grammar of the speaker's language. Chomsky's view of the role of intuition seems unacceptable in an empirical science: one frequently encounters hypotheses about the grammar of English, *e.g.*, which appear to be based on wishful thinking about the well-formedness or ill-formedness of various unusual word-sequences, but, given Chomsky's methodology, there is no possibility of refuting such hypotheses. Fortunately, once we appreciate the relevance for linguistics of the standard notion of the *ceteris paribus* clause, we can discard Chomsky's unempirical data-base without having simultaneously to give up the first- and second-order linguistic theories which Chomsky and his colleagues have constructed on that base.[2]

The research programme of linguistics aims to discover one first-order theory for each human language, together with a single second-order linguistic theory. (As in other sciences, in practice many rival candidates for correct first-order theory of a given language, and for correct second-order linguistic theory, are being canvassed at any given moment.) Note that the relationship of second-order linguistic theory to first-order linguistic theory is not simply one of greater generality, if by calling a theory *T* more 'general' than a theory *T′* we mean that the set of potential falsifiers of *T′* is a proper subset of the set of potential falsifiers of *T*. The sets of potential falsifiers of a first-order and of a second-order linguistic theory will commonly be disjoint. For instance, a second-order linguistic theory which declares the prime-number language for any vocabulary not to be a possible human language does not thereby entail any statement about individual word-sequences; the occurrence of any prime-length sequence in any language is perfectly compatible with such a second-order theory. (Thus, the five-word sequence 'We all came home late' is well-formed in English, although English is not a prime-number language.) Rather, grammars act as observational theories for the second-order linguistic theorist, as theories about the workings of telescopes and chronometers are observational theories for an astronomer who uses such instruments to obtain data bearing on some astronomical hypothesis.

---

[1] See Chomsky [1957], [1965], and other works; his [1961] is especially relevant for the point at issue here.

[2] See ch. 4 of my [1975] for more detailed discussion.

As in other sciences, the requirement that theories must not conflict with observation does not suffice to determine either unique grammars of individual languages or a unique general theory of language, and *a priori* criteria of theory-choice must be used to choose between alternative unrefuted theories at either level. Thus, we prefer strong theories: since a grammar is refuted by the utterance of a word-sequence which it forbids, the strongest grammar will be the one which permits fewest sequences. If strength were the sole criterion, then the best grammar for any language *L* might seem to be simply a finite list of the sequences which the linguist had heard *L*-speakers utter; but the additional criterion of simplicity will justify the choice of a grammar which uses generalised rules to define a set including the observed sentences together with other sentences which happen not to have been observed. (This situation is quite parallel to that described by Goodman, *op. cit.*) Similarly, the strongest second-order linguistic theory will be the one that defines the narrowest class of sets of word-sequences; but the criterion of simplicity will require the general theory to define a class of sets of word-sequences intensionally, rather than merely to list each of the several thousand natural languages for which data are available.

For Chomsky, the purpose of producing first-order linguistic theories—grammars—is as a means towards the goal of a good second-order linguistic theory. Chomsky holds that the correct general linguistic theory is stronger than one might expect *a priori*: in his terms, there exists a rich system of 'linguistic universals', or limitations on the diversity of natural languages. Chomsky further claims that one can infer various psychological and philosophical theses about human nature from the existence of these universals: he uses the linguistic universals to argue for a position which he labels variously 'nativism', 'rationalism', or 'mentalism'. However, although these claims are interesting in themselves, they are not relevant to the subject of this paper.

It is not even necessary, for present purposes, that the reader accept the validity of linguistics as a scientific discipline having the methodological structure outlined here. There are several reasons why one might be unwilling to do so. Thus, one might be unconvinced by my suggestion that the *ceteris paribus* clause in its standard role permits us to preserve Chomsky's theoretical findings while replacing his intuitionistic data-base with objective data about utterances. (If grammars are based on intuitions as Chomsky claims, then it will no longer be true that the strongest grammar is the grammar generating fewest sentences—since we make intuitive judgments of ill-formedness as well as of well-formedness—and much of the argumentation in the body of this paper will consequently lose its force.) Or, one might feel that the fact that language is an aspect of the behaviour of intelligent beings renders falsifiable predictions about utterances impossible in principle, in the same way that, as Popper says (*e.g.* in the preface of his [1957]), falsifiable predictions about the evolution of human knowledge are impossible in principle. I do not agree with these objections (on the latter point see ch. 3 of my [1975]); but, for the purposes of this article, it is enough if the reader accepts the logical possibility that there could be a science of some subject with the structure I have outlined, including separate theories describing individual examples of some class of phenomena together with a general theory describing the class exemplified by those individuals. Linguistics provides a concrete exemplification of the problem about theory-choice which I wish to take up, provided my account of its methodological structure is correct; but that problem is in principle independent of particular examples.

The problem has to do with the fact that, in a 'two-level' subject such as I have described, preferred theories at either level may be incompatible with preferred theories at the other level. The linguist who constructs a second-order theory tests its predictions against the individual languages he happens to know about, and his knowledge of these individual languages is commonly 'intuitive' in the unobjectionable sense, and thus taken for granted rather than thought of as tentative or provisional. But in reality the correct theory of any individual language is of course just as much an open question as the correct second-order linguistic theory; and indeed the same linguist may well change hats, so that on one day he is, say, an English grammarian, forming novel hypotheses about the correct theory of English, while on another day he is a general linguist testing novel hypotheses about the correct second-order linguistic theory against English as described by a theory which he does not question. There is nothing unusual in the fact that general linguistics relies on observational theories: all sciences do. The trouble is that, at least in linguistics, there tends to be an inverse relationship between 'goodness' of theories at the two levels, so that choosing a better theory at either level necessitates choosing a worse theory at the other level; and it is not clear which level should take precedence.

Let me give a hypothetical example. (The example is not implausible, and was suggested by a consideration of some facts about Classical Chinese; but any genuine example would introduce many irrelevant complexities.) Suppose that we possess a large enough corpus of writings in Atlantic (the language spoken on Atlantis before its destruction) to give us a rather good knowledge of that language: although none of us have that fluency in Atlantic which comes only with prolonged immersion in a living language, nevertheless many of us can read Atlantic literature for pleasure, can correct pupils' exercises in Atlantic composition, and so forth. Like Basque, Atlantic does not seem to belong to any of the known language-families; but in almost all ways Atlantic is a 'normal' language, in the sense that it differs from other known languages only in respects in which they differ among themselves. There is one notable exception. Atlantic has no subordinate clauses; or at least, none are attested in any of the Atlantic documents that have been preserved. In situations where an Englishman might write, say, 'He who laughs last laughs longest', or 'I know that you are angry', an Atlantean will write something that would translate literally as, say, 'One man laughs last; that man laughs longest', or 'You are angry; I know it', respectively. An English sentence containing subordinate clauses can normally be paraphrased rather accurately by a series of simple clauses; in Atlantic we find only simple clauses.

Yet every other language known to us (let us suppose) does have subordinate clauses. Should we hypothesise that Atlantic permitted subordinate clauses which happen never to occur in the large but finite corpus we have inherited; or should we say that Atlantic is exceptional and genuinely did not permit such constructions?

My description of the hypothetical situation assumes that the notion of 'subordinate clause' can be defined independently of the grammar of particular languages (otherwise it would make no sense to ask whether all languages permit them); and this is indeed so. Informally, a subordinate clause is a constituent part of a sentence which possesses the same internal structure as a complete sentence: it is easy to see that this description applies to the relative clause in 'He *who laughs last* laughs longest' and to the object complement in 'I know that *you are angry*'. The informal definition is far from watertight. For one thing, under that definition many natural languages, *e.g.* German, would lack subordinate clauses (in German verbs occur in second position in main clauses but finally in what are normally called 'subordinate clauses'); for another thing, the informal definition would count as subordinate clauses many word-sequences which would not normally be described as such (*e.g.* 'John came' is an English sentence but is not a subordinate clause in 'Hard on the heels of *John came* Peter'). But the informal definition can easily be made precise in the vocabulary of current second-order linguistic theorising.[1] Thus, a language permits subordination if the base component of its grammar contains at least one rule in which the initial symbol occurs to the right of the arrow (let us call such a rule a 'subordinating rule'); any constituent of a sentence dominated by an initial symbol which was introduced by a subordinating rule is a subordinate clause. The formal definition counts as subordinate clauses just those word-sequences which would traditionally be viewed as such. The vocabulary of that definition is part of the 'hard core' (Lakatos [1970], p. 133) of the modern general theory of language, so that, although we are discussing choice among second-order theories, it is quite reasonable to postulate that the question whether Atlantic contained subordinate clauses is a meaningful question with respect to any of the second-order linguistic theories under consideration.

I can find nothing intrinsically implausible in the imaginary situation I have described. We might even be able to construct hypothetical Atlantic sentences containing subordinate clauses, while having no way of knowing whether a living Atlantean would have accepted such sentences as perfectly normal or reacted to them with the baffled stare with which we greet English 'word-salad'.

It is true that the hypothetical problem depends crucially on Atlantic being a dead language; if it were a living language we could simply go on eliciting more sentences, knowing that, if subordinate clauses were grammatical, one would be sure to turn up eventually (a language cannot contain constructions which are never used, otherwise children would have no way of learning the constructions and they would therefore cease to be part of the language). But criteria for theory choice are supposed to tell us which of rival theories to prefer, given a finite quantity of observations:

---

[1] For a concise exposition of current second-order linguistic theory, see *e.g.* Kimball [1973].

the scientist is never in the position of choosing between theories in the light of infinitely numerous observations, even though there may be infinitely many potential observations to be made. The general problem of how to choose between grammars of Atlantic is surely not affected by the fact that the data-supply in this case was limited by the eruption of Thera rather than by the linguist's patience.

One might argue that Atlantic lacked subordinate clauses as follows: if subordination was grammatical, then the probability of finding no examples in a large corpus would be vanishingly small; we do find no examples, therefore the hypothesis that subordinate clauses were grammatical is almost certainly false. But this argument embodies an assumption about equiprobability which we have no right to make. One might quite reasonably suggest that Atlantic *did* have subordinate clauses, but that the Atlantean canon of literary aesthetics valued short, uncomplicated, punchy sentences, and that the many subordinate clauses that occurred in spoken Atlantic never showed up in writing for that reason. The data *might* decide between the alternative grammars of Atlantic indirectly (if, for instance, written Atlantic sentences were in fact often long and syntactically involuted, then the argument from probabilities might have more force); but it might equally well be that the data were neutral and the decision could only be made, if at all, in terms of *a priori* criteria for theory choice.

Yet the central criterion for theory choice, namely strength, gives contradictory results. Since all natural languages but Atlantic are known to permit subordinate clauses, and no natural language is *known* to exclude them, a second-order linguistic theory will be stronger if it asserts that all natural languages include subordinate clauses than if it asserts that some natural languages do and others do not include subordinate clauses. Therefore the criterion of strength applied at the level of general linguistic theory leads us to choose a theory which, among other things, implies that Atlantic does have subordinate clauses. On the other hand, the grammar of Atlantic will be stronger if it forbids subordinate clauses than if it permits them. Which should take precedence, strength of second-order linguistic theory or strength of Atlantic grammar?

My own inclination as a linguist would be to give precedence to strength of the grammar, and to say that Atlantic has no subordinate clauses. But I know the origin of this inclination, and it has little methodological basis. Like Chomsky, I am interested primarily in general linguistic theory and only secondarily in grammars of individual languages. Since my aim is to establish a strong general linguistic theory, I feel an obligation to pay due attention to apparent refutations of second-order hypotheses: there is no point in holding a strong theory if one ignores *prima facie* refutations.

Since lack of subordinate clauses in Atlantic would falsify an attractive second-order hypothesis, I feel bound to give the evidence against Atlantic subordinate clauses the benefit of the doubt.

But it is not clear that correct scientific methodology bids one give one's opponents the benefit of the doubt in this rather British way; and, in any case, were my reasons for being interested in language other than they are, the same logic would lead me to the opposite opinion about Atlantic. Thus, suppose I were a philologist with Atlantic as my specialisation, and suppose I were interested in general linguistics only in so far as it might be able to shed light on my studies of Atlantic philology. I might, in particular, be specially interested in the vexed question whether the portions of the *Timaeus* which are preserved in Atlantic translation were translated by a native Atlantean or by a Greek immigrant: the question is relevant for deciding to what extent philosophical thought had taken root in Atlantic civilisation. Now I shall hope to establish a maximally strong grammar of Atlantic: in particular, if I feel justified in inferring from the absence of subordinate clauses in original Atlantic literature that Atlantic lacked subordinate clauses (*i.e.* that native speakers of Atlantic would never speak or write sentences containing subordinate clauses), and if I nevertheless discover a handful of subordinate clauses in the Atlantic version of the *Timaeus*, than I shall be able to conclude that these are errors perpetrated by a non-native-speaker of Atlantic, so that the translation must have been produced by an immigrant. But now my inclination to give the other man the benefit of the doubt will prompt me to concede that, after all, the general linguists have produced good evidence that all natural languages have subordinate clauses, so I must regretfully admit that Atlantic probably had them too; and I can therefore draw no conclusion about the translator of the *Timaeus*.

Surely, however, we expect the philosophy of science to provide us with criteria for deciding what is the most reasonable belief given certain data, irrespective of any motives we might have for trying to reach a belief? The question 'Did Atlantic have relative clauses?', like the question 'Does Pluto have moons?', needs no context to make it a reasonable one; we are entitled to a less evasive answer from the philosophy of science than 'Why do you ask?'[1]

It may be suggested that my instinctive inclination to give precedence to strength of the grammar is not in fact merely a function of my greater personal interest in general linguistic theory than in individual languages, but rather has solid methodological justification. I know of two arguments purporting to show this.

---

[1] Levi [1967] argues that scientific theory-choice involves a calculation of the utility, in terms of "relief from agnosticism", of acquiring alternative beliefs; but Levi's principles do not imply specific utility-distributions for specific cases of theory-choice.

In the first place, it might be claimed that while the first-order hypothesis 'Atlantic lacked subordinate clauses' is falsifiable and therefore empirical, the second-order hypothesis 'All natural languages permit subordinate clauses' is of the wrong logical form to be empirical. The latter hypothesis might be regarded as an abbreviation of the statement 'For each natural language, some sentence of that language contains a subordinate clause': statements containing both universal and existential quantifiers can be neither conclusively verified nor conclusively falsified (unless the range of one of the variables quantified over is a finite set that can be exhaustively checked), and such statements are therefore metaphysical rather than empirical (*cf.* Watkins [1958]). However, it seems perverse to translate the hypothesis 'All natural languages permit subordinate clauses' as suggested, since there is no way to ascertain whether or not a given sentence of a language *L* contains a subordinate clause *unless we have a first-order theory* (either intuitive or explicit) *of L.* (Clearly there is no way of telling whether or not an isolated sentence in an otherwise unknown language contains a subordinate clause: no proper part of the sentence can be identical to the sentence itself.) Therefore a more reasonable expansion of the second-order hypothesis would be 'For each natural language, some rule in the grammar of that language will be a subordinating rule'. The grammar of any natural language will comprise a small finite set of rules, so this statement is conclusively falsifiable and thus empirical in form.

The other argument would be that to maintain the second-order hypothesis in the face of the Atlantic evidence is to render it unempirical in practice, even if it is empirical in form. The first-order hypothesis that Atlantic lacks subordinate clauses, it could be claimed, is unquestionably empirical: it is falsified as soon as we observe a sentence produced by an Atlantean which is not generated by our current grammar of Atlantic, but which can most simply be handled by adding a subordinating rule to that grammar. On the other hand, if facts such as those postulated to be the case for Atlantic are not accepted as falsifying the second-order hypothesis that all natural languages have subordinate clauses, then it is not clear that anything *will* be accepted as falsifying that hypothesis, and its apparent strength will be an illusion. The choice is not between two strong but incompatible hypotheses (the objection runs), but rather between a genuinely strong hypothesis about Atlantic grammar and an apparently strong but in fact unfalsifiable general linguistic theory: clearly we must prefer the former.

However, the second-order hypothesis would not necessarily become wholly unempirical, just because we chose to maintain it in the face of the Atlantic counter-evidence. The hypothesis makes predictions about the several thousand existing natural languages; it will still be empirical if we allow it to be falsified only by counter-evidence from several different languages.

It is true that such an interpretation of the second-order hypothesis, while leaving it empirical, would render it somewhat weaker. But, in the first place, the first-order hypothesis that Atlantic lacks subordination likewise can hardly be refuted by a single example: if we found just one Atlantic sentence containing a constituent of sentential form it would be most rational to account for that sentence as a *sui generis* idiom like English 'How do you do?', and we would not be justified in adding a subordinating rule to the grammar unless we found several different examples of subordinate constructions. And, more important (since the respective strengths of the first- and second-order hypotheses are not commensurable), if we take the line that strength of the observational theory must be given absolute precedence over strength of the higher-order theory, then we shall come up with highly counter-intuitive findings over and over again in describing any natural language. Suppose, for instance, that a linguist writing a grammar of German has observed the word-sequences 'Er heißt Kurt', 'Er heißt Werner', and various similar sentences, but happens not to have heard the sequence 'Er heißt Hans'. His general linguistic theory predicts that if some string containing a proper name is well-formed in a natural language, then the string that results from the substitution of another proper name will also be well-formed in that language (assuming one makes any morphological adjustments that may be necessary if the language is an inflecting one, like Latin): the general theory will have to be weakened if he wishes to permit 'Er heißt Hans' to be ill-formed in German. But the grammar of German will be stronger *if* this sequence is ill-formed; and there is no logical contradiction in supposing the sequence to be ill-formed—perhaps Germans always say 'Sein Name ist Hans' instead, in the case of that particular name. If strength of the observational theory takes absolute precedence, then the conclusion must be that 'Er heißt Hans' is ill-formed in German. But who would in practice be willing to accept such a conclusion on such evidence?

Goodman (and others) suggest that a criterion of simplicity must be allowed to interact with the criterion of strength in deciding between theories. It is unclear what kind of 'trade-off' relation between strength and simplicity should be assumed when the two criteria conflict. (Goodman [1961] seems to suggest that simplicity takes absolute precedence over strength; but this can hardly be right, since, in any field of science, the hypothesis 'Anything can happen' is likely to be simpler than all stronger alternatives.) In any case, though, simplicity looks unlikely to solve the kind of case under discussion. It would appear that a grammar of Atlantic will be simpler if it lacks subordinating rules than if it includes them, while a second-order linguistic theory will be simpler if it makes no provision for exceptions to the law that natural languages permit subordination. In other words, it seems that the criterion of simplicity merely reinforces the criterion of strength, in this case.

In view of current developments in linguistics, some linguists might dispute this. Part of the 'hard core' of the dominant, Chomskyan general-linguistic research programme asserts that grammars of natural languages will consist of a context-free 'base component' together with a set of 'structure-dependent transformations': loosely speaking, the base component states the range of semantic functions available in a language, while the transformations specify the syntactic devices used to express those functions in various circumstances. Many linguists advocate a so-called Universal Base Hypothesis (see *e.g.* Kimball [1973], p. 122), according to which the base component is the same for all grammars and only the transformations differ.[1] If languages other than Atlantic provide no refutation of the Universal Base Hypothesis, then the question whether Atlantic does or does not permit subordination might crucially decide whether that hypothesis can be maintained. If Atlantic has subordinate clauses, thus permitting the Universal Base Hypothesis to be maintained, then the universal base will need only to be stated once for all in the second-order theory, not separately in the grammar of each language; but, if Atlantic lacks subordinate clauses, then (since other languages have them) the Universal Base Hypothesis must be false and the grammar of Atlantic will need to contain an explicitly-stated base. Therefore, it might be argued, the grammar of Atlantic will be simpler (though less strong) if Atlantic is held to include subordinate clauses.

Again, however, the argument cuts both ways. If the Universal Base Hypothesis is incorrect, then the second-order theory need only state that grammars of natural languages incorporate some context-free base and some set of transformations; while, if the Universal Base Hypothesis is correct, then the second-order theory will have to contain an explicit statement of the universal base. So, on the assumption that Atlantic subordinate clauses are crucial for the Universal Base Hypothesis, the second-order theory will be simpler if Atlantic lacks subordinate clauses, even though the grammar of Atlantic will be simpler if Atlantic contains subordinate clauses. The Universal Base Hypothesis is currently an open question, so that we do not in fact know which way the simplicity criterion would push us when applied to the hypothetical data of Atlantic; but it seems likely that it would push us in one direction while we considered alternative grammars for Atlantic, and in the opposite direction while we considered alternative second-order linguistic theories.

---

[1] The Universal Base Hypothesis is condemned as vacuous *e.g.* by Peters and Ritchie [1969], but *cf.* Sampson [1973].

To shift the discussion from the evaluation of theories to the evaluation of Lakatosian research programmes will not solve our dilemma. In general linguistics, it is certainly true that we are faced not with one or a small set of alternative second-order theories but rather with a range of evolving theory-series, which frequently split into distinct branches as a group of scholars suggest some modification to what was previously treated as "hard core" and then develop their new ideas in order to meet objections and account for further data. A similar pattern appears in the work of grammarians of English, and (provided their data-base was reasonably large) would no doubt also occur among grammarians of Atlantic. Both the general linguist and the grammarians of individual languages aim (or should aim) to promote progression and avoid degeneration in their respective research programmes. But, if grammars are observational theories for the general linguist, then developments in a research programme at one level will entail compensatory developments in research programmes at the other level, and, in cases such as that discussed, developments which tend to progression will have to be compensated for by developments which tend to degeneration: Lakatos's methodology makes no prescription for this kind of situation.[1]

It does follow from Lakatos's work that isolated cases of theory choice, such as the one discussed, are less important than earlier philosophers of science believed for determining whether the scholar(s) making the choice are rational scientists. Lakatos shows that it takes a whole series of theory choices to answer that question. But it seems to lie in the nature of linguistics that problems which are similar in principle to that of subordinate clauses in Atlantic will occur repeatedly in practice. The discussion of 'Er heißt Hans', above, suggests that, in many cases where we happen not to have observed an utterance of some sentence which would be well-formed if the language in question was regular, we will be faced with the choice of treating the sentence as well-formed (and therefore adopting a less than maximally strong grammar), or treating it as ill-formed (and weakening the second-order theory by re-organising it to allow languages to display a novel kind of irregularity).

---

[1] Thus Fred d'Agostino suggests in Lakatosian vein that we should provisionally accept the strong second-order linguistic theory although that forces us to adopt a relatively weak grammar of Atlantic; we must tolerate the latter point as an 'anomaly' for the second-order theory, and anomalies of this kind will stimulate us to search for some novel, as yet unthought-of second-order theory which will combine the best of both worlds (in that it will allow some grammar for Atlantic which is at least as contentful as the stronger of the two grammars currently under consideration, while itself being at least as contentful as the stronger of the two second-order theories discussed above). If such a theory can be found, then what is now reluctantly tolerated as an 'anomaly' for the stronger current second-order theory will be seen with hindsight as a 'refutation' of it. I accept the point that one *prima facie* falsification should not force us to abandon an otherwise well-supported theory; but, instead of provisionally accepting the stronger second-order theory and treating the weak Atlantic grammar as an anomaly for that theory, I do not see why it would not be equally rational to accept provisionally the stronger Atlantic grammar and treat the weak second-order theory as an anomaly for that grammar. Lakatos's view of methodology seems neutral as between these alternatives.

For philosophers such as Kuhn and Feyerabend, of course, the dilemma might appear to be merely a *reductio ad absurdum* of the premiss that there exist rational criteria for scientific theory choice. I do not accept this as a solution to my problem, since I find Kuhn's and Feyerabend's accounts quite inadequate, if interpreted as prescriptions as to how the scientist should operate (they may be more acceptable as descriptions of how the scientist does operate). Since Kuhn and Feyerabend suggest that he who convinces is right, the statement that I find them unconvincing seems a more telling objection to their views than it would be to the views of others.

Even writers who advocate relatively formalised theory-choice criteria of course recognise that theories may be non-comparable. For Popper [1934], if the set of potential falsifiers for theory *A* overlaps with the set of potential falsifiers for theory *B* but neither set is included in the other, then theory *A* and theory *B* are incommensurable. Our case is not like that: we are choosing simultaneously between two commensurable theories of Atlantic and between two commensurable general linguistic theories, and it happens that the preferred choice at one level is incompatible with the preferred choice at the other level. If we treat the situation as a choice between the stronger grammar of Atlantic and the stronger second-order theory, then the sets of potential falsifiers in the two cases will not even overlap (the grammar is falsified by the occurrence of utterances, the second-order theory by the existence of languages): Popper does not discuss the problem of deciding in such cases, since, normally, theories whose falsifier-sets are disjoint will not be incompatible.

The reason why problems of the kind discussed in this paper are particularly obtrusive in linguistics has to do with the two-level nature of linguistic theorising, whereby a given language is treated on the one hand as a specific example of the genus 'human language', and on the other hand as a general system to which individual sentences must conform. The subject-matter of the 'hard sciences', physics and chemistry, does not appear to contain similar structuring. One well-established science which might seem to parallel linguistics more closely in this respect is biology: a given species is on the one hand a particular example of the class of biological species, and on the other hand a pattern which is itself exemplified by individuals of that species. However, the nearest analogue in biology to the second-order theory in linguistics would presumably be the theory of evolution, and Darwinian evolution is notoriously not a predictive theory (Scriven [1959]); it explains in general how species come into being, but says nothing about which particular future developments will occur and does not delimit a range of possible future species. If the notion of 'strength' fails to apply to the higher-level theory in biology, it is understandable that conflicts between theories at the two levels of biology will not occur. However, although linguistics seems to be unusual in the regularity with which preferred higher-order theories are incompatible with preferred observational theories, presumably such conflicts may be expected to arise sporadically in any science in which a predictive theory depends on observational theories which are themselves predictive.[1]

The relationship between linguistics and the philosophy of science has until recently been a one-way affair: linguistics has had much to learn from the philosophy of science, but nothing to teach it. In the present case, I have no answer to the problem I have posed; but it is perhaps a sign that linguistics has come of age, if it generates a problem whose solution will form a contribution to the general philosophy of science.

University of Lancaster

---

[1] Noretta Koertge suggests that a possible parallel to my example might arise in physics in connection with the generalisation (which we might regard as a 'second-order law of physics') that all ('first-order') laws of force are inverse-square laws. As I understand it, a problem analogous to mine would arise if one found *prima facie* evidence that some novel physical force varied with distance according to a different formula: would it be better to abandon the second-order law, or to preserve it and abandon the first-order law (which would presumably involve treating the *prima facie* regularity as a merely coincidental patterning in an essentially random phenomenon)? My own knowledge of physics is too scanty to permit me to follow this idea further.

REFERENCES

Chomsky, A. N. [1957]: *Syntactic Structures*. Mouton.

Chomsky, A. N. [1961]: 'Some methodological remarks on generative grammar'. *Word* 17, 219–39.

Chomsky, A. N. [1965]: *Aspects of the Theory of Syntax*. MIT Press.

Feyerabend, P. K. [1970]: 'Against method', in M. Radner and S. Winokur (eds.): *Minnesota Studies in the Philosophy of Science*, **4**, University of Minnesota Press.

Goodman, N. [1961]: 'Safety, strength, simplicity'. *Philosophy of Science*, **28**, 150–1.

Kimball, J. P. [1973]: *The Formal Theory of Grammar*. Prentice-Hall.

Kuhn, T. S. [1962]: *The Structure of Scientific Revolutions*. University of Chicago Press.

Lakatos, I. [1970]: 'Falsification and the methodology of scientific research programmes' in I. Lakatos and A. Musgrave (eds.): *Criticism and the Growth of Knowledge*, Cambridge University Press.

Brit. J. Phil. Sci. **26** (1975), 319–330 *Printed in Great Britain*

A Critique of some Formal Theories of Meaning

by PAUL ERNEST

1 *Logical Preliminaries.*
2 *Definitions of Sense and Intension for L.*
3 *Extensionality.*
4 *Anomalies in the Treatment of the Logical Theory.*
5 *Anomalies in the Treatment of Mathematical Theories.*
6 *An Anomaly in the Treatment of Empirical Theories.*
7 *Dependence on L.*
8 *Equality of Sense and Intensions.*
9 *Definitions of Reference and Extension for L.*
10 *Innovations in the Definitions of Reference and Extension.*
11 *Relations between Intension (or Sense) and Extension (or Reference).*
12 *Definitions of Meaning for L.*
13 *A Proposed Explication of Intension for L.*

A significant trend in twentieth-century philosophy of mathematics has been the formal explication of previously vague concepts. Notions such as proof, effective procedure, logical truth and logical consequence have been explicated in a satisfactory way. Two recent formal theories of meaning are perhaps best viewed in the light of this tradition. These theories of meaning are due to Mario Bunge ([1972]) and Charles Castonguay ([1972]). Bunge has formulated a theory of meaning for the expressions of logically reconstructed empirical theories. Castonguay is more concerned with mathematics and its foundations, and he applies his theory of meaning to the expressions which occur in these fields.

Both theories rest on a dual analysis of the concept of meaning. Bunge analyses meaning into sense and reference and Castonguay analyses meaning into intension and extension. In each case, the sense or intension component of meaning is explicated by means of concepts supposedly from proof theory and the remaining component of reference or extension is dealt with in terms of model theory.

The languages most frequently used in the logical reconstruction of both empirical and mathematical theories are first order predicate languages

Received 24 *March*, 1975

with equality. For this reason, and since both theories of meaning are formulated mainly for first order languages, the exposition of the theories of meaning that follows is given for first order predicate language with equality.

1 *Logical Preliminaries*

Let L be some fixed but arbitrary first order predicate language with equality. Thus the language L has an unspecified but, in general, non-empty set of extralogical letters in addition to the usual logical connectives, variable letters and equality symbol. Suppose that L is provided with the rules of formation of L-formulas, definitions of a *free variable* and an *L-sentence*, and a classical deductive calculus.

The concepts of an *L-theorem* and that of an L-formula being the *deductive consequence* of a set of L-formulas are defined in the usual way. The formal and semantic relations of entailment are denoted by $\vdash$ and $\vDash$ respectively.

Let the lower case letters x, y, z, . . . and f, g, h, . . ., and the capital letters F, G, H, . . . be metalinguistic variables ranging over individual variables of L, L-formulas and sets of L-formulas, respectively. For any set F of L-formulas let $Cn(F)$ be the set of all deductive consequences of F under the axioms and rules of inference of L. A set F of L-formulas is said to be an L-theory, if F is a set of L-sentences which contains all L-sentences which are deductive consequences of F.

Following Castonguay, the ***rank*** of an L-formula f, ***rank***(f) is defined to be the number of distinct free individual variables occurring in f.

With this terminology the following definitions may be formulated:

2 *Definitions of Sense and Intension for L*

For any set of L-sentences F and any L-formula f, Bunge defines $Sen(f)$, *the sense of f taken with respect to F* by $Sen\ f = \{g \in Cn(F)\colon \vdash f \rightarrow g \text{ or } \vdash g \rightarrow f\}$.

Castonguay defines $Int(f)$, ***the intension of f taken with respect to F*** by $Int(f) = \{g\colon \boldsymbol{rank}(g) = \boldsymbol{rank}(f) \text{ and } F \vdash f \rightarrow g\}$.

These definitions constitute proposed explicata of the informal notions of sense and intension. The success of an explication of an intuitive concept depends on how faithfully what are deemed to be the essential properties of the intuitive concept are preserved by the explication. This criterion is central to the following critical examination of the two definitions.

3 *Extensionality*

Sense and intension are usually understood not to be reducible to reference and extension, respectively. The two authors concur with this requirement of irreducibility, both for any proposed explications of these concepts and, in particular, for their own explications (see for example Bunge [1972] p. 319 and Castonguay [1972] p. 6). The claim of Bunge that his explication of sense is not reducible to his explication of reference is justified, but only because his novel definition of reference does not coincide with the common notion of reference, which is identifiable with extension. It remains to be seen whether the explicatum of sense is reducible to extension.

Reiterating: it seems natural to require that definitions of sense and intension should be non-extensional, that is not reducible to extension. It further seems likely that the two authors have used proof theoretic concepts in their definitions in the hope of making them non-extensional. This choice seems well justified, for proof theory is, in general, non-extensional. Proof theory is constructive and concerned with particular proofs of formal theorems or with particular deductions of formal consequences from sets of sentences, and these objects of concern have a non-extensional character. For example an L-theorem f will have many different proofs in L and each of these proofs could perhaps be seen as relating to a different sense of f. At any rate, the particular characteristics of any one proof of f are non-extensional, although the unadorned fact that f is provable clearly is referential since it is equivalent to the logical validity of f.

Unfortunately, the authors have omitted to use any of the concepts of proof theory which have just been characterised as non-extensional and for this reason they have failed to give non-extensional explications of the concepts of sense and intension. This may be demonstrated as follows. Let alternative definitions of the sense and intension of an L-formula f relative to F be given by

$$\mathit{Sense}(f) = \{g : F \vDash g \text{ and } (\vDash f \to g \text{ or } \vDash g \to f)\}, \text{ and}$$

$$\mathit{Intension}(f) = \{g : \mathit{rank}(g) = \mathit{rank}(f) \text{ and } F \vDash f \to g\}.$$

These definitions of sense and intension are given entirely in the semantics of L and thus, being given in terms of purely extensional concepts, they are themselves extensional. By the Completeness Theorem for L, the expressions $F \vDash g$, $\vDash f \to g$, $\vDash g \to f$ and $F \vDash f \to g$ are respectively equivalent to $g \in Cn(F)$, $\vdash f \to g$, $\vdash g \to f$ and $F \vdash f \to g$. Thus $\mathit{Sense}(f) = \mathit{Sen}(f)$ and $\mathit{Intension}(f) = \mathit{Int}(f)$. The definitions of sense and intension of Bunge and Castonguay may be given in exclusively extensional terms, and thus fail to be non-extensional in any respect. This is a serious failing and in virtue of it neither definition can hope to adequately explicate our full, non-extensional, informal concepts of sense and intension.

As pointed out above, the senses and intensions of formulas are relative to the choice of a particular theory, F. In the following three sections, the consequences of choosing F to be the purely logical theory, a mathematical theory, and finally an empirical theory are examined. Attention has been restricted to L-theories for the following reason. If G is a set of L-sentences it is easy to see that the sense (intension) of an L-formula f taken relative to G is identical with the sense (intension) of f taken relative to the L-theory generated by G.

4 *Anomalies in the Treatment of the Logical Theory*

Let F be the purely logical theory; then F is the L-theory generated by the empty set of L-sentences. In this case, the definitions of sense and intension give rise to certain anomalous results. If f is any L-formula, the sense of f relative to F according to the definition of Bunge is $Sen(f) = Cn(F)$, the set of all theorems of L, since all L-theorems are deducible in L from any L-formula. Thus all formulas of L have the same sense relative to F. In effect, Bunge's explication of sense fails to distinguish between any formulas of L at all, in a purely logical context. This failure is perhaps not very significant, since Bunge's theory of meaning is intended for use in the contexts of reconstructed empirical theories, and F is not such a theory.

In contrast, the greatest power of discrimination is achieved by Castonguay's definition of intension precisely when F is a purely logical theory. Two L-formulas of the same rank have the same intension when their formal equivalence is provable from F. When F consists of theorems of L alone, this equivalence must be proven from the axioms and rules of inference of L without the aid of any extralogical postulates. Thus fewer L-formulas have the same intension relative to F than they would have relative to any theory of L containing extralogical postulates.

Even under the favourable circumstances of the choice of F as a purely logical theory, the definition of intension fails to make essential distinctions. For example, all of the infinite number of sentences which are theorems of L have the same intension. Many of these sentences formally express statements which have differences of meaning crucial to mathematics and the foundations of mathematics. These differences of meaning are obliterated by the explication of intension, in the present purely logical context and consequently for any other choice of F as well.

5 *Anomalies in the Treatment of Mathematical Theories*

Suppose F is a mathematical theory, that is F is generated by a set of mathematical postulates in L. If f and g are two deductively equivalent

L-formulas, that is $\vdash f \leftrightarrow g$, then relative to F,

$$Sen(f) = Sen(g)$$

In defence of this conflation of senses it might be argued that deductively equivalent L-formulas have differences of sense which are empirically insignificant. Thus sense as defined by Bunge, trivialises non-empirical distinctions. This definition does, however, assign distinct senses to L-theorems and L-formulas which are derivable from F, but which are not L-theorems in their own right. This is as it should be.

Intension, as defined by Castonguay, does not make such distinctions. If f and g are any two L-formulas of the same rank whose equivalence is deducible from F (that is $F \vdash f \leftrightarrow g$) then $Int(f) = Int(g)$.

This result seems indefensible, since intension has been defined specifically for mathematical theories. In particular, if we take F to be Zermelo–Fraenkel Set Theory, the most recondite sentence of set theory deducible from F has the same intension as the most (set theoretically) trivial sentence which is an L-theorem. For example, relative to this choice of F,

$$Int(V = L \rightarrow GCH) = Int(\forall x[x = x])$$

where $V = L$ is the Axiom of Constructibility and GCH is the Generalised Continuum Hypothesis. Furthermore, this intension is shared with an infinite number of other sentences of L. The absurdity of these results is self-evident.

From the above and previous criticism, it seems unlikely Castonguay's definition of intension can be of much use in mathematics.

6 *An Anomaly in the Treatment of Empirical Theories*

Let F be the formalisation in L of an empirical theory. Bunge distinguishes empirical theories from the types of theories we have previously considered as follows. Like a mathematical theory, an empirical theory consists of the deductive closure of a set of postulates which are not all logical theorems. Associated with this theory are certain semantical rules which specify the empirical interpretation of certain of the terms in the theory. These semantical rules may be taken as restrictions on the class of intended models of the theory, and for the purposes of determining senses these semantical rules may be disregarded. Suppose that F has a finite axiomatisation in L, and let the L-sentence h be the conjunction of this finite set of axioms. Attempts to reconstruct empirical theories might well lead to finite axiomatisations, so Bunge's definition of sense should deal adequately with the sense of h. The sense of h relative to F is

$$Sen(h) = Cn(F).$$

Unfortunately, if g is any counter-theorem of L (that is $\vdash \neg g$), we also have

$$Sen(g) = Cn(F)$$

since g, being inconsistent, has any L-formula as a deductive consequence. In particular,

$$Sen(h) = Sen(h \,\&\, \neg\, h).$$

Thus the definition of sense has the anomalous consequence that the sense of any L-formula finitely axiomatising an empirical theory is identical with the sense of any contradictory statement in L, taken relative to the given theory. Bunge's definition of sense thus suffers from a serious defect: it is not satisfactorily applicable in a wide range of empirical contexts, those where an empirical theory is finitely axiomatisable.

Next, consider a defect both definitions suffer from, and a possible remedy.

7 *Dependence on L*

The definitions of both sense and intension have the unfortunate property of being highly dependent on the arbitrary initial choice of the language L.

Choose a first order language L_0 which differs from L in its set of extralogical letters. Then any formula of both L and L_0, relative to a set of formulas of both L and L_0, has a sense (intension) in L_0 which differs by an infinite number of formulas from the sense (intension) of the formula in L. To illustrate this more precisely, let f and F be a formula and a set of formulas respectively, of L. Let L_1 and L_2 be two first order predicate languages with equality which, in addition to the set of all extralogical letters of L, have each a distinct set of new extralogical letters. If g is a formula of L_1 but not of L, then each of the L_1-formulas $g \to g$, $g \to (g \to g)$, $g \to (g \to (g \to g))$, . . . belongs to $Sen(f)$ in L_1 but not to $Sen(f)$ in L.

Further, relative to F $Sen(f)$ in L_1 and $Sen(f)$ in L_2 are incomparable with respect to set theoretic inclusion.

The strong dependence of the definitions of sense and intension on the particular language used, is a defect. No such dependence seems to be required by the informal notions of sense and intension, and many interesting concepts, such as logical validity, satisfiability and provability have language independent explications.

This difficulty can be avoided by restricting the formulas which belong to the sense or intension of an L-formula f to those in the language with the smallest set of extralogical letters in which F and f remain expressible. Thus the definitions of sense and intension are given for the poorest language in which F and f are expressible.

8 *Equality of Sense and Intensions*

Neither the definition of sense nor that of intension allows the effective determination of whether two formulas f and g of L have the same sense or intension, relative to some set of L-formulas F. If we are fortunate enough to establish that $\vdash f \leftrightarrow g$ or $\vdash f \rightarrow g$, then we know that $Sen(f) = Sen(g)$ or $Sen(g) \subseteq Sen(f)$. To establish similar relations among $Int(f)$ and $Int(g)$ we must determine whether $F \vdash f \leftrightarrow g$ or $F \vdash f \rightarrow g$. In none of these cases is there a uniform effective method, so we may never know how $Sen(f)$ and $Sen(g)$ relate to each other. Except in the case where f and g differ in rank, the same is true of intensions. In this case, $Int(f) \cap Int(g) = \phi$. This latter result in itself is not entirely acceptable, for however much intuitive meaning two formulas seem to share, if they differ in rank their intensions are disjoint. Consider the two L-formulas $x = x \,\&\, y = y$ and $x = x$; intuitively these formulas have a great deal of intension in common yet, according to Castonguay, their intensions are disjoint (assuming x and y are distinct variables).

In summary, it has been shown that both definitions fail to capture the full objective content of the intuitive concepts of sense and intension. Both definitions are extensional, highly language dependent and give rise to anomalies.

Next, definitions of reference and extension are considered.

9 *Definitions of Reference and Extension for L*

Let M be an L-structure. If f is formula of L, Bunge defines $Ref(f)$ *the reference of f relative to M* to be the union of the fields of the denotations under M of all the extralogical letters occurring in f. If F is an empirical theory, as was mentioned earlier, there are certain semantic rules associated with F. These rules specify the denotations or classes of denotations permissible for some of the extralogical letters of L. In effect, these semantic rules delimit a class of L-structures which empirically interpret F and provide L-formulas with their intended references.

Let U^i designate the ith Cartesian product of U and put $U^0 = \cup \, \{U^i : i \in N\}$. Suppose f is an L-formula of rank n. Then Castonguay defines *the extension of f relative to M* by:

$$Ext(f) = \{u \in U^n : f \text{ is satisfied by } u \text{ in } M\}.$$

If f is a sentence of L then f is of rank zero and either f is satisfied in M by all u in U^0 and $ext(f) = U^0$, or f is satisfied in M by no u in U^0 and $ext(f) = \phi$. Thus U^0 and ϕ represent the truth values Truth and Falsity, respectively.

These definitions of reference and extension are largely based on the classical theory of models. It is the difference between these definitions and analogous model-theoretic concepts which is the next consideration.

10 *Innovations in the Definitions of Reference and Extension*

It is clear that whatever concept Bunge has explicated with his definition of reference, it is not the Fregean concept of reference. Bunge defines reference so that it is independent of truth values. This can be seen as follows: for any L-formula f and relative to any L-structure M, $Ref(f) = Ref(\neg f)$. Thus truth values cannot be determined by this concept of reference. This in accordance with Bunge's intentions ([1972], p. 318), for in his theory of meaning, truth values belonging to the real interval [0, 1] are assigned to formulas in an unspecified way entirely separate from the assignment of references. The purpose of defining reference as Bunge does is, roughly speaking, to provide a means of indicating those objects in the universes of the intended models which a formula 'speaks about' or refers to. I am not convinced that this concept of reference is of much use.

The sole innovation in Castonguay's definition lies in his taking the idiosyncratically defined set U^0 and the set ϕ to be the truth values Truth and Falsity, respectively. This innovation is intended to concretise the truth values within the Universe of Sets and to give extension a superficially smooth definition. However, the result is that the truth values lose their universal character since U^0 as defined by Castonguay depends on the choice of L-structure. Castonguay would have done better to stick to the usual definition of U^0 as $\{\phi\}$, since then his truth values would correspond to the set theoretic definitions of 1 and 0. In this case, the definition of extension would coincide with the usual Tarskian notion.

Bunge and Castonguay both lay heavy emphasis on the joint articulation of the components of their pairs of definitions, which is examined in the following.

11 *Relations Between Intension (or Sense) and Extension (or Reference)*

It has been argued that the relation of inclusion between the extensions (or references) of two formulas is entailed by the converse relation of inclusion between the intensions (or senses) of the same two formulas. On the informal level such considerations are as vague as the concepts involved. On the more exact level of first order predicate languages, both Bunge and Castonguay subscribe to such a principle. It is easy to see that Bunge's definitions do not satisfy such a principle. However, since Bunge's concept of reference is not the usual one, there is no reason to expect such a principle to be satisfied. Bunge proposes an alternative concept of ***horizontal intension*** which would relate to reference in the above manner; since this concept is not made definite, it has not been discussed in this article.

Castonguay, however, stresses that such a relation between intensions and extensions is necessary to vindicate his explications ([1972], p. 13). Specifically, given any *L*-formulas f and g, Castonguay claims to prove that, relative to suitable F and M

$$\text{If } \mathit{Int}(f) \subseteq \mathit{Int}(g), \text{ then } \mathit{Ext}(g) \subseteq \mathit{Ext}(f). \tag{*}$$

A convincing proof of (*) would show that however unacceptable Castonguay's definitions might be, at least they relate to each other in what is, perhaps, an intuitively acceptable manner. A consequence of (*) is

$$\text{If } \mathit{Int}(f) = \mathit{Int}(g), \text{ then } \mathit{Ext}(f) = \mathit{Ext}(g). \tag{**}$$

It is surely non-controversial to claim that for an explication of intension to be adequate, it must satisfy this condition.

Having seen that Castonguay requires (*) to hold, the motivation behind the use of the concept of rank in the definitions of intension becomes clear. Since *L*-formulas of differing non-zero ranks have disjoint extensions, for there to be any hope of proving (*) and (**) these same *L*-formulas must have disjoint intensions.

Unfortunately, Castonguay's alleged proof of (*) ([1972], p. 21) is defective. Whilst the order of free variables is immaterial in considerations involving intensions, in the case of extensions this order is crucial. Thus, two *L*-formulas which have the same intension and free variables, but whose variables occur in different orders may well have different extensions. For example, consider the two *L*-formulas $x = x \,\&\, y = c$ and $y = c \,\&\, x = x$ in the case where L contains a constant letter c. Relative to any consistent set of *L*-sentences F, these two formulas have the same intension. However, given any model M of F with a non singleton universe, it is easy to see that the extensions of the two formulas differ. This contradicts (**), and hence (*). This technical defect may be remedied by changing the definition of rank. If f is an *L*-formula which contains the distinct free variables $x_1, \ldots, x_n$ in the listed order, define $\text{rank}(f) = \langle x_1, \ldots, x_n \rangle$. Alternatively, define the rank of f to be the n-tuple of subscripts of the variables $x_1, \ldots, x_n$. In either case, it is easy to see that (*) and (**) will be satisfied.

Bunge and Castonguay complete their theories of meaning with definitions of the concept of meaning in terms of the previously defined concepts.

12 *Definitions of Meaning for L*

If F is a set of sentences of L and M is an L-structure, Bunge defines the meaning of an L-formula f relative to F and M to be

$$\langle Sen(f), Ref(f)\rangle.$$

Castonguay similarly defines the meaning of f relative to F and M to be $\langle Int(f), Ext(f)\rangle$, but requires in addition that M models F.

Bunge's explication of meaning gives rise to the following anomaly. Reconsider the case where an L-theory F is finitely axiomatised by a single L-formula h. It is easy to see that relative to F and to any L-structure M, h and $h \,\&\, \neg h$ have exactly the same meaning according to Bunge.[1]

The strategy of defining meaning as an ordered pair is an attractive one, and no criticism of it is intended. It is rather the definitions of the components of meaning, particularly the definitions of sense and intension which have been criticised. The definitions of sense and intension are criticised for failing to satisfy several conditions laid down for any adequate explications of sense and intension. These adequacy conditions for explications of intension or for its approximate synonym sense, do not seem to me to be overly stringent. It is possible to explicate intension so that these conditions are fulfilled. To demonstrate this I include a proposal for an explication of intension for L which seems to fulfil the conditions of adequacy suggested in this article.

13 *A Proposed Explication of Intension for L*

The intension of an L-sentence f is determined by the syntactic structure of f and the meaning of the logical symbols of f prior to any interpretation of the extralogical letters occurring in f. In consequence, a proposed explication of the intensions of L-sentences should employ a method which fundamentally depends on the syntactic structure of L-sentences and on the meaning of the logical symbols of L without depending on any other property of L-sentences. The method of semantic proof tableaux[2] is a suitable candidate for such a method since it proceeds by the analysis of the syntactic structure of L-formulas by means of rules which reflect the meaning of the logical symbols of L.

A proof tableau for an L-sentence f consists of the result of a systematic search for a proof of f. It has the form of an inverted tree structure, possibly of infinite depth, which is constructed during an attempt to prove f by deriving a formal contradiction from $\neg f$. There are many ways of specifying a complete tableaux procedure by which a unique proof tableau is constructed for an L-sentence f. Suppose that one of these procedures is chosen for its economy in proving provable L-sentences, so that any L-sentence f has a canonical tableau $T(f)$. My proposal is that the intension of an L-sentence f has $T(f)$ as its explicatum. Thus the intension of an L-sentence f is the uniquely specified method of the verification of f.

---

[1] It should be mentioned that the account of Bunge's theory of meaning is taken from from the brief and rather schematic account in [1972]. A fuller exposition has been published, see M. Bunge [1975*a*] and [1975*b*]. Unfortunately, these works were unavailable at the time when this article was written.

[2] For an exposition of semantic tableaux see, for example, Smullyan [1968].

Having sketched the proposed explication of intension for L, it remains to be seen whether it satisfies the adequacy conditions. These conditions are:

(1) The identity of intensions should be effectively determinable.
(2) The explication should be non-extensional, but intension should determine extension.
(3) The explication should not depend on the particular first-order language used.
(4) The intensions of sentences taken in the contexts of formalised theories should not give rise to anomalies.

My tentative proposals meet these requirements as follows:

(1) Two L-sentences with identical canonical tableaux are themselves identical. Thus the determination of the identity of intensions of L-sentences is trivially effective.
(2) Any two different L-sentences will have different formal intensions whether they have the same extension or not. Furthermore, extensionally equivalent L-sentences generally have canonical tableaux which differ *significantly* in many respects. The formal intension of an L-sentence f determines the extension of f for $T(f)$ is a proof of f if and only if f is valid.
(3) The canonical tableau $T(f)$ of an L-sentence f is identical for any first-order language L_0 in which f is expressible. This is because $T(f)$ depends only on the extralogical symbols which occur in f.
(4) The approach sketched above can be extended (see Smullyan [1968]) to assign intensions to expressions of the form $F \vdash f$, where f is an L-sentence and F is a set of L-sentences. If F and G are two sets of L-sentences representing different axiomatisations of the same consistent mathematical or empirical theory, the expressions $F \vdash f$ and $G \vdash f$ will have different intensions (canonical tableaux) for any L-sentence f. If f and g are distinct L-sentences $F \vdash f$ and $F \vdash g$ will

have different intensions, so the proposed treatment of intensions relative to theories avoids the conflation of intensions for which the definitions of Bunge and Castonguay are criticised. For any consistent set of L-sentences F and any L-sentence f the expressions $F \vdash f$ and $F \vdash \neg f$ have different formal intensions, irrespective of whether f axiomatises F or not.

The proposal I have given for the explication of intensions for L is tentative and unfinished, but I hope that it suggests that the adequacy conditions for explications of intension which I have listed can be fulfilled.

Chelsea College, University of London

REFERENCES

BUNGE, M. [1972]: 'A Program for the Semantics of Science', *Journal of Philosophical Logic I*, 1972. Reprinted in M. Bunge (*ed.*): 'Exact Philosophy—Problems, Tools and Goals', Dordrecht: D. Reidel, 1973.

BUNGE, M. [1975*a*]: *Treatise on Basic Philosophy*, 1: *Sense and Reference*. Dordrecht: D. Reidel.

BUNGE, M. [1975*b*]: *Treatise on Basic Philosophy*, 2: *Interpretation and Truth*. Dordrecht: D. Reidel.

CASTONGUAY, C. [1972]: *Meaning and Existence in Mathematics*, Berlin: Springer, 1972.

SMULLYAN, R. M. [1968]: *First-Order Logic*, Berlin: Springer, 1968.

Review Articles

BEYOND TRUTH AND FALSEHOOD*

This rather idiosyncratic book contains a number of theses or themes, interdependent and intertwined in various ways. Some of these themes are openly, conspicuously and deliberately personal, and we are plainly invited to see them as such. The personal aspect has a number of elements: a dialogue with the late Imre Lakatos, a settlement of scores with Sir Karl Popper and his followers, and a personal manifesto/self-portrait by Paul Feyerabend.

The volume is dedicated to Imre Lakatos, described as 'fellow-anarchist', and the introductory note specifies that 'every wicked phrase' in the book was written in anticipation of 'an even more wicked reply' from Lakatos. This may well give the impression that Imre Lakatos shared the author's 'anarchist' views and style, at least in some measure. For the record, and in fairness to the memory of Imre Lakatos, it is important to state that this fostered impression bears no relation to the truth. What is the case is that, in his private life, Lakatos was much given to agreeable clowning, which gave pleasure to his friends, and such clowning *in private life* may perhaps resemble, to some extent, the clowning found *in print* and within hard covers in Feyerabend's book. But in his work, his writings and lectures, Lakatos observed the highest standards of rigour, lucidity and responsibility. Not only did he observe such standards, he was, as Feyerabend himself reports in another part of the volume, deeply and systematically concerned with the maintenance of such standards in the face of current sloppy trends (anarchistic in Feyerabend's sense). Lakatos's élitism in educational policy, which he had the opportunity of supporting both under a Communist and under a Western regime, sprang precisely from this concern. It did not arise from some over-all right-wing position; he continued to be left-wing on economic issues, and did not seem unduly interested in hierarchy and authority, one way or the other, in society at large. It was the threat to *intellectual* standards which aroused him to strong feeling, whether it came from Communist egalitarianism or from Western permissiveness. By contrast, Feyerabend's book, if taken at all seriously, can only be a charter for the abandonment of all and every standard. It is inconceivable that Lakatos would have allowed himself to be hailed as a 'fellow anarchist' in this sense. The dialogue which did indeed go on between him and the author concerned certain specific issues about scientific method.

The claim that Lakatos's position was 'anarchism in disguise' is totally unfounded. It might perhaps be true that Lakatos has not solved the problem which arises from the divergence between the real practice of science on the one hand, and the picture of it offered by writers on method on the other. But it is totally irresponsible to present this failure (if such it was) as a tacit or camouflaged admission that the problem is insoluble, and that consequently 'anything goes'. For it is *this* argument which is at the heart of Feyerabend's book. There is nothing to show that Lakatos either embraced this conclusion or was logically obliged to do so.

---

* Review of P. K. Feyerabend [1975]: *Against Method.* London: New Left Books. £5.75. Pp. 339.

The relationship to Popper is different. Popper and his followers are attacked on various grounds, over and above holding erroneous views:

> Popper's philosophy, which some people would like to lay on us as the one and only humanitarian rationalism in existence today, is but a pale reflection of Mill's *On Liberty*. It is . . . much more formalistic and elitist, and quite devoid of concern for individual happiness . . . We can understand its peculiarities when we consider . . . the unrelenting puritanism of its author (and of most of his followers), and when we remember the influence of Harriet Taylor on Mill's life and on his philosophy. There is no Harriet Taylor in Popper's life (p. 48, *n*).

This is a piquant passage. My own liberalism is such that I hold that even puritans are not debarred from truth (some of my best friends are puritans) and that even professors of philosophy are free to refrain from having mistresses, if their inclinations are so eccentric. This should not lay them open to censure or render their views suspect. But perhaps I push liberalism too far.

But there is a serious element in all this. Feyerabend is not the only man of talent, closely associated with the Popperite movement, who after some time displays a need to lash out against it. W. W. Bartley III and Imre Lakatos himself are other names which spring to mind. It is difficult not to suppose that excessive demands for conformity and involvement, made by the master, are at the root of such subsequent revulsions. It is ironical that the foremost proponent of critical philosophy should have formed a movement so intolerant of criticism that it drives its own members to such outbursts. Those of us who admire Popperian ideas without having been involved in the movement, can only deplore the way in which, logically or not, the ideas come to be tarnished by such an atmosphere. Feyerabend's suggestion that this could be avoided by diversifying one's emotional involvements has its plausibility, though people should perhaps be free to choose the form which the diversification takes. In fact, not all single-minded people are necessarily intolerant, and even the possession of a harem does not necessarily make a ruler liberal. So Feyerabend's characteristically Viennese cure for the closed society can hardly be the whole answer.

Feyerabend's attitude to himself, which is affectionate and admiring, is the third markedly personal theme in the book. It is relevent in as far as the author repeatedly makes plain that this book is not to be judged by the normal conventions, such as are generally taken for granted by the authors, readers and reviewers of books. It is a *happening*. It makes its own convention or rather it suspends a lot of them, and switches them on and off, and above all it invokes the meta-convention that no conventions are binding. The game is played by the author's own rules, which moreover he is free to change as he wishes. This naturally makes it somewhat difficult for the reader to assess or argue with the author's ideas. Just what kind of happening this book is meant to be, and by what criteria if any it could be judged according to its author, we can only guess at in the light of what he says about himself and his endeavour.

> . . . Puritanical dedication and seriousness . . . I detest . . . I now prefer . . . the term *Dadaism* . . . A Dadaist is utterly unimpressed by any serious enterprise . . . is convinced that a worthwhile life will arise only when we start taking things *lightly* . . . is prepared to initiate joyful experiments even in those domains where change and experiments seem to be out of the question . . . I hope . . . the reader will remember me as a flippant Dadaist . . . (p. 21, *n*.).
>
> . . . the epistemological anarchist has no compunction to defend the most trite, or the most outrageous statement . . . His aims remain stable, or change as a result of argument, or of boredom, or of a conversion experience, or to impress a mistress, and so on . . . The one thing he opposes positively and absolutely are universal standards, universal laws, universal ideas such as 'Truth', 'Reason', 'Justice', 'Love' . . . (p. 189).

We can hardly complain that we have not been given the self-imposed terms of reference of the work. More of this hereafter.

In fact, the book contains a number of related theses and attitudes, some of which are of interest. In addition, there are secondary theses elaborated by way of support or defence for the main ones, a kind of intellectual spin-off from the main argument. What I would select as the primary theses, the backbone of the book, forms a kind of series or progression. It is worthwhile enumerating the vertebrae in this spinal column, summed up in my own words:

(1) The actual history of science shows that the real advances of knowledge contradict all available methodologies.

(2) This shows that all methodological doctrines or principles are false.

(3) This is so not because the existing theories of scientific method aren't good enough, but because *no* principles of method could ever be good enough.

(4) As it is difficult to distinguish very general principles of method from criteria of truth, it follows not merely that we cannot distinguish between good and bad methods, but equally that we cannot distinguish between good and bad substantive theories.

(5) Hence let a million flowers bloom, for at least two reasons (which to me seem contradictory): we cannot tell the good ones from the bad ones, *and* the proliferation aids the emergence of good ones.

(6) In our intellectual and cultural and other activities, there are many aims more important than truth, and for this reason, in addition to the above, we should not bother with truth and the uniqueness of truth. (This argument would seem to me otiose if we cannot in any case identify truth).

(7) All this being so, we should just fool around as the fancy takes us. I, Paul Feyerabend, am fooling and clowning for all I'm worth, at this very moment, and all the time.

This seems to me the central plot, though as I say, there are side-arguments and spin-offs which also deserve consideration. The elements in this spinal column are related to each other in various ways which give the book its organic unity. (1) is clearly the nucleus from which the whole thing grew. This is the area of Feyerabend's professional expertise, and where a claim that he is engaged in sustained debate with serious scholars (even if he himself also disclaims seriousness) can be upheld. One way to treat this book would be to concentrate on the extensive parts which argue this viewpoint in connection with the history of modern physics, and to ignore the rest as off-duty clowning, which had, through temperamental idiosyncracy, and encouraged by the spirit of the times, been allowed to intrude within the covers of the book. Such a procedure would be reasonable, and perhaps the most sensible one to adopt. I shall not adopt it, partly because it could only be usefully carried out by a historian and/or philosopher of physics, and partly because the rest of the book does seem to me of some interest, even if only as a phenomenon rather than a serious position, and I do feel competent to discuss it.

(2), (3), (4) and (5) are developments of (1), through various forms of strengthening or extension. Where (1) has a certain plausibility, these extensions or extrapolations lack it, but nevertheless they deserve some discussion. The most obvious criticism is of the step from (1) to (2) and (3): if existing methodologies do not do justice to the actual process of the advancement of knowledge, why can we not improve on these methodologies, and how could we know that it cannot in principle be done? (Perhaps I may at this point state my own view of the matter. I doubt whether any methodology can fully capture the complexity of the transition from one theory to another, or from one style or paradigm to another. The transition will always hinge on the *content* of the theory or paradigm, which of course must be ignored by an abstract and general theory of method, meant to be applicable impartially to all such advances. On the other hand, theories of scientific method *can* do the important job of singling out worlds amenable to science from those which are not. This is important because not all worlds or thought-styles are amenable to science.)

Feyerabend tries to establish (3) with the help of some additional tributary considerations which feed into the mainstream of the argument. History is far too complex to be encompassed within the simple bounds of any theory of method; and secondly, *reasons* are never sufficient to effect changes, of opinions or of anything else, and the inevitable *causes* cannot be incorporated in methodologies which can only specify reasons. On the first point, he invokes the authority of Lenin. All this seems to me to prove too much. There is a methodology for solving quadratic equations, which many schoolboys and others frequently implement. Of course, each real occasion of such implementation is more complex, *and* can also be interpreted causally. But this well known general consideration can be invoked to exclude the possibility of *any* rational or norm-governed action. Are we to conclude that no-one ever really solved quadratic equations? This problem is in no way specific to the question of whether scientific methodology actually governs the history of science. Feyerabend makes no cross-reference to the relevant philosophical history of this much wider problem, but proceeds, here as elsewhere, as if he'd just discovered the point, and as if it were *specifically* relevant. His awareness of the relevent parts of general philosophy seems highly limited. Elsewhere, on the other hand, he blithely carries on as if no such impediment to human fulfilment existed, and reasons could liberate us. In brief, (3) neither follows from (1), nor is cogently, or even consistently, supported by additional arguments.

By contrast, at the next stage, *if* (3) be granted, then (4) does follow, and, given some not very contentious additional premisses, or tastes, so does (5). If there are no methods and hence no criteria, if follows that all substantive propositions are equally good or bad. If this is so, then in turn we only need a preference for activity as against passivity, to end up with anarchism rather than with scepticism. Given our alleged inability to distinguish between good and bad ideas, we can refrain from all of them (scepticism), or capriciously indulge in any that take our fancy (Feyerabendian anarchism). It would seem a matter of taste. Feyerabend notices that his argument could equally lead to scepticism, and it is in fact at many points indistinguishable from it. We seem to owe the preference for anarchism to his temperamental exuberance:

> Epistemological anarchism ... differs from scepticism ... While the sceptic ... regards every view as equally good, or equally bad ... the epistemological anarchist has no compunction to defend the most trite, or the most outrageous statement (p. 189).

Given (5), *i.e.* given anarchism/scepticism (choose your option, and presumably anarchist freedom includes scepticism as one of its own freely available internal options), the further transition to (6) also seems to me quite natural. Moreover, (6) has a certain independent plausibility, and many thinkers have reached it, or some position similar to it, without having come to it by the questionable path trodden by Feyerabend. Moreover, (6) can be supported by some additional considerations which are also present in the book. These are the arguments from the human condition, and from diverse cultural styles of life. (These are my terms, not Feyeraband's.)

The argument from the human condition runs as follows (when summed up in my own words): those of us who are honest, know that our intellectual life is a mess. Epistemologies pretend that each of us inhabits a coherent world, and that we reached it for good reasons, and know it to be rationally preferable to any other world open to us. In fact, we inhabit an incoherent chaos, we don't quite know how we acquired all the bric-à-brac that comes to make it up, though we *do* know that some bits of it we acquired improperly or even shamefully, by means we could not defend; and we also opportunistically keep all kinds of alternative visions and strategies in reserve, and hope to get by somehow. Coherent, clear and rigid visions are fraudulent façades, as are strong and consistent moral characters; and ideologies which try to bully us into pretending that things are otherwise, that good and manifest reasons are available to justify their own advertised options, commit violence against us and against the truth of the matter. All this is so, and constitutes the grain of truth in Feyerabend's more general anarchism (as opposed to his more specific thesis about divergence between methodology and the real history of science). This strengthens his case: he reminds us in a salutary manner of the way things really are, and of the amount of pretence involved in the official theories. But it is one thing to say that in a complex and difficult world, almost anything might contain some truth, and therefore it behoves us to be humble; and quite another to say that anything goes and therefore we should be arrogant.

Then there is the argument from culture and thought-styles. The point is familiar. There are many rival cultures and values in this world, often competing within single societies and within the breasts of single individuals. The choice between such styles or cultures is difficult, and certainly there are no easy or self-evident choices. But each culture or style of life tends to carry with it, and internally validate, its own world-picture and its own cognitive norms. Hence cultural relativism, or liberalism as between cultures, also forces us into cognitive relativism. This view is reinforced by the notorious difficulty or impossibility of proving basic premisses of world-pictures (a consideration which also strengthens the argument from the human condition).

Thus amongst the more plausible of Feyerabend's premisses are the view of the irrelevance of methodologies to the actual progress of science, and the idea that diverse cultures or values are incommensurate, and that the superiority of one of them (*e.g.* science) should not too easily be taken for granted. He uses *each* of them to support the view that anything goes, that an extreme form of relativism is valid. My own view is that the two problems in some measure cancel each other out in the sense that the main point of epistemologies is precisely that they provide a way of making inter-cultural comparisons. Epistemologies are indeed, I suspect, irrelevant to the manner in which *specific* cognitive advances are reached. Their real use lies in helping us make inter-cultural comparisons. A culture which subjects its cognitive capital to testing by arbiters *who are not under its own control*, seems to me superior to one which does not do so. Epistemological theories thus do give us some insight in how to choose between whole styles of thought. But I do not expect them to apply mechanically to individual discoveries or advances, even *within* the culture which, as a whole, satisfies this or that epistemology.

All this is worth discussing. My complaint here is that, quite contrary to all appearances, Feyerabend, far from being too extreme (which is how he delights to present himself), is not nearly extreme enough; or, to put it another way, that his extremism fails through its frivolity. The trouble is, he reaches the problem of how we choose between different styles of life and their associated cognitive norms not, as the problem deserves, because it is important, relevant and intriguing; he reaches it by a process of wilfully exaggerating the implications of his conclusions about the relation of methodologies to the real life of science. Even then, he does not really reach the *problem* or discuss it properly, he merely reaches the anarchistic or relativistic solution of it, and he does that far more for the sake of trying to extract every ounce of *épatement* from it, than because he is seriously concerned with the question to which it might be a (regrettable) answer. The question is not properly discussed, nor the implications of the answer seriously explored. It is all just part of the clowning, an exaggeration of certain views, pushed to an extreme to achieve an effect or to avert boredom or to satisfy a jaded intellectual palate. As he tells us, his

> . . . aims . . . change as a result of argument, or boredom, or of a conversion experience, or to impress a mistress . . . (p. 189).

I only hope that she was more impressed than I am. There are some gullible mistresses about, but that I knew anyway.

One might however, object that the argument from the existence of rival cultural styles of knowing, between which it is difficult to arbitrate, only justifies *cultural* relativism (with whole cultures as the units to which truth is 'relative'); whereas Feyerabend's anarchism is relativistic in a more extreme form, going beyond even individual relativism and reaching to the justification of the autonomy of every mood, every caprice, and every individual—which must never be eliminated simply because some rival and incompatible belief has better support . . .

However, if this extreme individual-occasion anarchism/relativism of (5) and (6) be granted, then (7), an attitude rather than a belief, ('I'm clowning, and clowning is best') *does* seem to follow. In a world as described in (5) and (6), clowning does seem to make more sense than any alternative. But if clowning is the only permissible strategy ('anything goes'), as Feyerabend delights to repeat, then of course, all the logical inadequacies in the previous build-up, not to mention the lack of seriousness, receive *ex post* clearance. It was all clowning, it proved that only clowning is the right method, and as no reasoning can fall below the standards of clowning, then by the very standards which we have now established, the reasoning which led us here was entirely sound. The circle is complete. The final step of the argument conclusively, by its own norms, underwrites all the steps which led up to it. Of course, some other clown might clown differently, and so might *this* clown when the mood takes him; but then, Feyerabend does not deny this. He says of himself:

> To be a true Dadaist, one must also be an anti-Dadaist.
> His (the epistemological anarchist's) favourite pastime is to confuse rationalists by inventing compelling reasons for unreasonable doctrines. There is no view, however 'absurd' or 'immoral', he refuses to consider or act upon . . . The one thing he opposes positively . . . are universal standards . . . though he does not deny that it is often good policy to act as if such laws . . . existed . . . (p. 189).

Clearly Feyerabend has invented a game at which he cannot lose. Of course, it is not internally coherent: the conclusion that all ideas have equal right to our assent, is supported, over and over, by arguments which presuppose that on the contrary we do know some of them to be true; that we know, for instance, the truth of certain generalisations, such as that all scientific theories are contradicted by some facts, or that some scientific theories continue to be upheld even when falsified. But a conclusion which claims to show that the requirement of coherence, like all other requirements, is illegitimate, obviously cannot be refuted by the demonstration that the approach to it contained incoherent steps . . .

So shall we leave it at that?

No. There are things to be said about the inadequacy of Feyerabend's *Problemstellung*, about his solutions, about the sources of that solution, and about his style.

We are in effect presented with the choice between two possible worlds. Each of these worlds is attained and ratified by a distinct cognitive strategy. Each has its own epistemology as its Founding Charter. The first world is the one normally inhabited by most Western academics. It is a world which historically owes something to the jealous and exclusive Jehovah, to the Platonic ideal of unique mathematical truth, to the rigour of Roman law, and to the scripturalism of the Reformation. These traditions all converge on the crucial idea that truth is *unique*, and is surrounded by a vast sea of variegated error. The idea of *method* which accompanies this vision, is that of a narrow and perilous path which leads, and which alone leads, to cognitive salvation. Science as conventionally conceived has taken over this picture. *Extra scientiam nulla salus* is how Feyerabend wittily and appositely (p. 306) sums up this conception.

Within this conventional world, the practice of clowning is also known, and is governed by certain conventions. These conventions preclude us from subjecting a clown's pronouncements to the same kind of scrutiny and criticism as we should apply to serious contention. To violate Clown's Licence is to show oneself a boor.

Then there is another and Alternative World, so warmly commended by Feyerabend. This world is infinitely rich and manifold and, according to Feyerabend, contains human beings greatly superior, freer and more fulfilled, than those who reside in the normal world. The underlying epistemology of this other world, its key and entrance, is the principle—*anything goes.*

These are the *dramatis personae.* What is the plot? This, alas, depends on what the reader does. If he carefully examines—as I think he should—this fascinatingly radical challenge to our usual assumptions, the idea that that this Other World is indeed superior to the one we normally take for granted, and if in the course of doing so he dares raise some objections—he is scornfully dismissed as one who pedantically and meanly fails to observe a rule of *this* world, which exempts clowns from logical scrutiny . . .

If, however, he allows himself to be browbeaten by this into going along with the joke, he suddenly finds himself, not in the presence of a *jeu d'esprit,* but under summary judgment from a People's Court set up by the New Regime to root out all reactionary survivals from the Old World . . .

So clowning is, all at once, both a criticism-evading ploy within *this* world and one invoking its conventions, *and* the Founding Epistemology of *another* world which would supplant this one, and excoriate its citizens as slaves of a rigid and oppressive orthodoxy. It is all at once an argument-suspending evasion, and a warrant for condemnation of others.

Feyerabend's clowning is not total and unremitting, nor is his preaching of capricious incoherence self-exemplifying. Or perhaps one should say instead that he *is* consistent in self-exemplifying his secondary rejection of all consistency, which excludes even the consistency of inconsistency. (You can't win.) Anyway, contrary to his primary commendation of inconsistency, and consistently with his secondary denial of consistency even in inconsistency (this regress has the form of a kind of recursive wobble), he does argue with a measure of coherence. Let us consider for instance his handling of epistemology.

Perhaps an epistemology must do justice both to the good grounds for doubt *and* to at least the strong presumption that we do have some knowledge. Feyerabend does a good job on the first requirement, but is hopeless on the second. Epistomology has to explain, or explain away, the *differential* problem—that some bits or putative knowledge seem so much better than others. This occurs at both individual and at social level. Some cognitive traditions seem conspicuously more effective than others. Consider the individual and social problems in turn.

We are notoriously confident about certain predictions, such as that we shall die if we step out of the window of a 50-storey building. Feyerabend reports Lakatos as insistently invoking this argument. In fact, Lakatos liked using the argument against Popper's views as well as Feyerabend's:

> Where is the epistemological anarchist who out of sheer contrariness walks out the window of a 50-storey building? (p. 221).

Feyerabend thinks he can answer this in terms of irrational compulsions of the reluctant anarchist.

> He may readily admit that he is a coward, that he cannot control his fear, and that his fear keeps him from the windows ... What he *does* deny is that he can give reasons for his fear ... (pp. 221 and 222).

People who genuinely have fears which they also sincerely consider irrational, have no objection to being cured of those fears, if a painless cure is available. In fact, certain drugs are said to induce fearlessness in people in such circumstances. I am however, convinced (not irrationally, I believe) that the reluctant anarchist, when offered the appropriate drug on the 50th floor, will also firmly refuse it.

Except for the more modern psychological idiom, this is of course a new version of that aspect of Hume's theory which gave a psychologistic account of our causal beliefs, for lack of a rational one. Feyerabend presents his summary psychologism in a cavalier manner as a new discovery, without the least hint that it may possess either difficulties or a prior history.

Let us move from his failure to deal with the difference between our *individual* firm convictions and our genuine doubts, and move on to the difference between technologically effective and powerless *styles of thought*.

If we adopted his principle 'anything goes', would not the consequence be a collapse of that science and technology on which we have come to rely for affluence, and without which the world's swollen population would both starve and tear each other apart in a far more savage conflict than we endure at present? Interestingly, Feyerabend does not deny the dependence of technology on that kind of square science in which it is *not* the case that 'anything goes'. This is an enormously significant concession: most people, when asked to say why it is *not* true that 'anything goes', would invoke the supremely important pragmatic consideration that some thought styles lead to effective control over nature, and others do not. No epistemology which ignores this Big Divide can claim to have asked the right question. Feyerabend does ignore it in the mainstream of his argument, but, curiously, accepts it when dealing with the (for him) tangential matter of the social consequences of his own position:

> ... while an American can now choose the religion he likes, he is still not permitted to demand that his children learn magic rather than science at school ... The separation between state and church must therefore be complemented by the separation between state and science.
>
> We need not fear that such a separation will lead to a breakdown of technology. There will always be people who prefer being scientists to being the masters of their fate ... Greece developed ... because it could rely on the services of unwilling slaves. We shall develop ... with the help of numerous *willing* slaves in universities and laboratories who provide us with pills, gas, electricity, atom bombs, frozen dinners ... (pp. 299 and 300).

So, when it comes to securing these benefits, pills and all (the pills naturally come first on his list), there is a welcome admission that 'anything' evidently does *not* 'go'. With a coherent or coherence-recognising thinker, one might conclude that his epatant relativism is but skin deep: subjected to the acid test of practical effectiveness, relativism is abandoned. So perhaps it all boils down to the far less provocative and in effect quite mild contention, that there are aims other than truth, and that a scientific civilisation is in danger of ignoring or underestimating those other aims? A reasonable viewpoint, but hardly original, or deserving of all that Dadaist trumpet-blowing. In fact, it is a view often proposed by utterly square people . . . How ironic that a genteel platitude should lurk under all that apparent extravagance.

But note the social philosophy which accompanies it. It is a blatant recommendation of parasitism, Californian style. The opting out from cognitive and productive rationality is recommended on the basis of an open expectations that others will not do likewise, and that we can continue to benefit from their exertions. Though the helots are recognised as necessary for our comfort, we are at the same time invited to despise them. Arrogance and parasitism seem to lurk under the *gemuetlich*, all-permissive liberation.

The position is parasitic twice over. The 'anything goes' principle is recommended partly as a piece of self-confirming anarchism, but partly also as a consequence of a more interesting claim—that the proliferation of rival viewpoint aids the advancement of knowledge. But natural selection from amongst a host of contestants only works if there is, indeed, *selection*: it is not enough to say that the more contestants, the merrier. If there is no selection, the multiplication of contestants on its own will have no effect whatever. The selection will only work if people do *not* accept incoherence, pointless proliferation of viewpoints, and so forth. So the effectiveness of the recommendation hinges on it being ignored, just as the material basis of the epistemological anarchist hinges on others not being anarchists—though they only get despised for their pains.

His doctrines about violence are similarly Californian. In a pronouncement early in the book, he tells us that

> A Dadaist would not hurt a fly—let alone a human being.

One is glad to hear it, though the persistent references to the pleasure that is to be had by confusing poor souls ('rationalists') who argue in good faith, does make one wonder. One's fears are later confirmed:

> Violence is *necessary* to overcome the impediments erected by a well-organised society . . . and it is *beneficial* for the individual, for it releases one's energies and makes one realize the powers at one's disposal (p. 187).

'Violence', if the word means anything, involves imposing one's will on a reluctant person who happens to be weaker. (Vigorous exercise with inanimate objects is not violence.) So violence does indeed help one person to realise the powers at his disposal, whilst giving another one a good taste of impotence and humiliation. The mystique of violence is justified as a reaction to a "well-organised society", and not given any specified limits. It is conflated incoherently with a pacifist no-fly-hurting posture, and goes with cognitive/productive parasitism.

His remarks on education are in the same tradition. In as far as a coherent doctrine can be extracted from his remarks, his view seems to be that education should be anarchist and tentative, not prejudging such open issues as the relative merits of science and magic, but leaving choice to the individual. But, oddly enough, 'progressive' theories of education are abused, though contemporary 'progressive' educational theories come close to his kind of 'anarchism'. He notes that Lakatos was 'concerned with intellectual pollution', such as the

> empty verbiage full of strange and esoteric terms claims to express profound insights, 'experts' . . . tell us about our 'condition' and the means for improving it . . . (p. 217).

Not a bad characterisation of *Against Method*, in which a melange of truisms and extravagances (hedged by self-characterisation as deliberate provocation) is presented as a recipe for our liberation.

His views about the general traits of our social condition are in the same style. For instance, whilst contemporary liberal society is apparently, as we saw, so oppressive that if justifies uncircumscribed violence on part of the individual, *other* societies (generally far more oppressive) receive indiscriminate praise in this hysteria of protest. For instance

> primitive tribes . . . solve difficulty problems in ways which are still not quite understood . . . there existed a highly developed and internationally known astronomy in the old Stone Age, this astronomy was factually adequate *as well as* emotionally satisfying, *it solved both physical and social problems* . . . (p. 306).

Feyerabend appears to possess information about the emotional states and social organisation of old Stone Age man which has been withheld from more conventional anthropologists and archeologists, no doubt because they are not aware of the methodological principle that 'anything goes'. In similar spirit, political interference in science in China and the USSR is commended, though apparently it was abortive in the case of Lysenko—who, by some most arbitrary decision within Feyerabend's terms of reference, does not benefit from the principle that anything goes.

Distant obscurantisms and authoritarianisms benefit from the principle—only the mild local liberal discrimination between reason and fantasy, and between the rule of law and chaos, does not. The history of the past four centuries is summed up as

> the suppression of non-Western tribes by Western invaders (p. 299).

So much for the non-tribal civilisations of the Eastern world. The ulama, Brahmins and mandarins are all relegated to the status of tribesmen. The only difference from the usual picture is that the poison brought by the conquerors is not capitalism and colonialism, but science and rationalism. Perhaps all this nonsense is simply a case of indulging his self-proclaimed 'favourite pastime' which is to 'confuse rationalists by inventing compelling reasons for unreasonable doctrines'. Unfortunately, this rationalist at any rate finds himself, far from compelled—just embarrassed at seeing someone make such an exhibition of himself. Far from compelling anything, this stuff can only be saved by the catch-all defence that 'anything goes'.

Has he, to use the kind of colloquialism he enjoys, gone off his rocker? Or is it all an elaborate joke? No. It only falls back on being a joke when under criticism. It borrows its social attitudes from Californian dissent, and its motive from the need to lash out with maximum force against one specific academic closed society. This motive seems to drive the author to any position of supposed maximum outrage, in accordance with the well-known internal mechanics of tantrums, when the child looks round for the most potent verbal missile that may be to hand. The author appears so intoxicated with his own charm that he evidently thinks these antics are irresistible, and repeatedly nudges us to notice how liberated, daring, rich in ideas, and free of noxious constraints he is. Discussing a critic who dared note an inconsistency in his work, he writes

> . . . Machamer . . . raises the ghost of papers I wrote centuries ago (subjective time!) to combat something I wrote more recently. In this he is no doubt influenced by philosophers, having made some tiny discovery, come back to it again and again for want of anything new to say and who turn this failing—lack of ideas—into a supreme virtue, *viz.* consistency (p. 114).

No such weakness in our author, evidently: his ideas are not tiny, but big, and flow in so constant a stream that evidently we need not fear the pseudo-virtue of consistency. And look how truly big those ideas are: comparing primitive societies with scientific ones, in favour of the former of course, he notes

> true, there were no collective excursions to the moon, but single individuals . . . changed into animals and back into humans again . . . (p. 307).

Scepticism about this truly amazing achievement would no doubt indicate failure to appreciate the principle that anything goes, and enslavement to the rigid thought-styles imposed by rationalism. The clowning posture he derives from some tradition which teaches that light-hearted *Gemuetlichkeit* is the cure for the world's ills. The trouble is that clowning only has charm if it is good-natured, and has an element of humanity and humility. This clowning is persistently rasping, boastful, derisive and arrogant; its attitude to what is rejected is aggressive and holier-than-thou, and opponents are not allowed to benefit from the all-permissive anarchism; the frivolity contains a markedly sadistic streak, visible in the evident pleasure taken in trying (without success) to confuse and browbeat the 'rationalists', *i.e.* people who ask questions about knowledge in good faith. This is why what might otherwise seem a harmless piece of Californian-Viennese *Schmalz* leaves such a disagreeable taste in the mouth.

ERNEST GELLNER
London School of Economics

Brit. J. Phil. Sci. 26 (1975), 343–351 *Printed in Great Britain* 343

THE PRAGMATIC MAXIMUM[*]

The dual tendency to view the pragmatists collectively under the aegis of their 'ism' and pragmatism in turn as a rural philosophy of the American frontier has damaged the reputations of non-pragmatic thinkers as far back as Jonathan Edwards (1703–1758)[1] as well as obscured the originality and individuality of the pragmatists themselves. Despite the searching scholarship of such as Morton White, W. B. Gallie, Richard J. Bernstein, A. J. Ayer, and now Scheffler, the myth of American philosophical naivete during the eighteenth and nineteenth centuries persists by allusion and in reviews[2] if not among those who have bothered to read more than the predictable snippets so dear to the anthologists.

To be sure, post-Edwardsean philosophy in mid-century before the pragmatists was something of a sustained let down. The pre- and anti-Kantian 'Mental Philosophers' of Lockean and Common Sense inspiration, the majority of college teachers of philosophy between 1820 and 1870, wrote in the drab and eclectic style suitable to textbooks of the day and their conservative college congregations. Theirs was a terminal case of hardening of the categories complicated by conflicts between the old theology and the new science that Common Sense and piety could not solve.[3] But out of their struggles, however muted in their books, with Edwards's idealistic determinism (the original 'block universe') and the 'evidences' of introspection came the vision of a new science of the mind which influenced that peculiar blend of social science and philosophy most conspicuous, perhaps, in James's work but no small feature of that of Peirce, Mead and Dewey. This blend, at once so fecund and befuddling, at the point of bifurcation of philosophy and psychology supplies the critical *leitmotif* of Scheffler's study of the four main pragmatists. He views them as doing philosophy at the watershed of social science, sometimes in complementary fashion, sometimes confounding them, usually when attempting a *general* account of the relations among, for example, thought, habit, and action, inquiry and knowledge, truth and certainty, deliberation and choice, fact and value, thinking and symbolisation. A fascinating tapestry of good whole cloth emerges from Scheffler's compact scrutiny of the original arguments and accumulated vital commentary.

Scheffler aims 'to portray the major themes of the four main thinkers of pragmatism, with special reference to their theories of thought and knowledge' placing 'primary emphasis on the views of the individual thinkers themselves' (p. *ix*). He writes 'not as an advocate of pragmatism, or as an interpreter merely, but primarily as a sympathetic critic' (p. *x*). The *post mortem* atmosphere of the standard histories is displaced by that of live confrontation. Unsurprisingly, new strength comes through the old texts even as Scheffler relentlessly pursues their weaknesses in a series of alternating expository and critical chapters. Following a useful thematic introduction and short biographical introductions to the philosophers themselves, the book proceeds in this dialectical fashion keeping the spotlight on thought and knowledge while also sweeping in a variety of related issues from the arts, sciences, religion, and education. The entire discussion is amply supported by a refreshing selection of quotations, and the comparative 'bridge work' among the four philosophers is solidly constructed. The overall effect is like looking out over a complex terrain with a good map in hand. The following comments touch on only a few of the significant features of Scheffler's discussion of each of the four main pragmatists.

---

[*] Review of I. Scheffler [1974]: *Four Pragmatists*. London: Routledge & Kegan Paul. Pp. vii + 259. £4.75.

[1] Whose analysis of the language of choice and causation in *Freedom of the Will* (1754) anticipates many manoeuvres of twentieth-century analytic philosophy.

[2] C. K. Grant is the latest to indict himself along with C. D. Broad in the former's review of the present book, Grant [1975].

[3] The most widely used textbook was *The Elements of Mental Philosophy* (1834) by Thomas C. Upham (1799–1867). An elegant writer and expositor of the Common Sense philosophy, particularly as it applies to the emotions, Upham nevertheless totally misconstrued Kant, for instance, as advocating Understanding and Sensibility as two sources of two sorts of ideas.

Peirce's theories of thought and inquiry are pursued through examinations of his cosmology, critique of Descartes, analysis of scientific method, theory of meaning, and his educational and religious views. Clearly the most rigorous and systematic thinker of the group, Peirce is perhaps more to be applauded for his ability to see a problem than to solve it. Indeed the often hopeless complexity of his answers sometimes renders them less interesting as answers than as a measure of the profoundity of the complex questions he raises. Such is the case with that part of his cosmology which attempts to work out an evolutionary principle of scientific method.

As a scientist *and* logician, Peirce realised that science is no mere theory-out-of-fact construction, that laws and theories are always underdetermined by the available data with innumerable theories, some of them incompatible with one another, being capable of accounting for the same set of facts (p. 26). Accordingly, 'Law is *par excellence* the thing that wants a reason' (quoted on p. 27) so far as the justification of scientific method is concerned. In fact, Peirce was troubled by two major problems: first, that uniformity in nature as reflected in scientific law and theory is as much in need of explanation as particular events; and second, that explanation by subsumption of particular laws under more general ones is inadequate to that task because it leads to an infinite regress. Cosmology begins where Peirce attempts to halt the regress by replacing the inscrutable theory-out-of-fact account with his own equally inscrutable habit-out-of-chaos theory of cosmic growth.

Scheffler acknowledges Peirce's first, diagnostic point but challenges the second, and hence, the necessity of the latter's genetic account of universal law. After noting that 'Peirce in fact glides from the question of explaining particular laws to that of explaining regularity in general' (p. 33), Scheffler goes on to argue for a kind of 'provisionality' more in tune with Peirce's own doctrine of fallibism. 'Such provisionality means we must be prepared to revise our judgments as to what is, in fact, a law. It does not imply that the very notion of a law as expressing a universal relationship must be given up in favour of an evolving regularity' (p. 37). This is because 'Science . . . does not attempt to explain all regularities at once, but, at any given time, assumes some to explain others. This process is neither circular nor committed to dogmatism [of fixed, first principles]' (pp. 34–5). Neither of course does it involve an infinite regress, since the analysans of any one explanation may be the analysanda of the next in what might be described as a recusant pattern of inquiry.

Much more is involved on both Peirce and Scheffler's parts than these bare bones, but it has seemed to many, including Scheffler (p. 27) that here lies ***the*** inconsistency in Peirce's work, namely, a cosmogonic philosophy incompatible both with his theory of meaning and his theory of inquiry. Certainly it lies without the bounds of the operationist Pragmatic Maxim; but then so does most of science. Likewise, as an extra-scientific theory of the origins of law it lies without the bounds of an evolutionary theory of inquiry as well. But as a theory *within* science about the *development* of law, it falls within the purview of the latter as a further instance of inertial belief plagued by doubt in search of a reason. As Quine has recently observed, '. . . we can fully grant the truth of natural science and still raise the question, within natural science, how it is that man works up his command of that science from the limited impingements that are available to his sensory surfaces'.[1] Quine goes on to note that this is a question of 'empirical psychology' of evident philosophical interest—just the sort of mixed problem bag that Peirce could not resist. Such a proto-scientific construal of Peirce's cosmology renders it no less vague but far less 'metaphysical' in principle at least. The problem is not so much in the very idea of such an inquiry[2] as in Peirce's tendency to get off on the foot of first causes.

The discussion of Peirce's 'doubt-belief theory of inquiry' (p. 65 ff.) is particularly valuable. While faulting Peirce for confusing psychological and epistemological analyses of inquiry in the insistence that all inquiry begin in 'real and living doubt' (p. 69), Scheffler continues to probe for the deeper insight:

> One may . . . try to read . . . the theory of inquiry as *epistemology* rather than psychology. Its import then is not to *describe* the course of thinking, but rather to say, for example, that *proper* scientific thought always is directed to specific questions, that it proceeds always by taking certain assumptions for granted, and that it is to be evaluated by how well it satisfies the questions from which it arose. Read thus, it rejects the idea that there can be scientific investigation without any assumptions whatever, as well as the idea that required assumptions must be absolutely indubitable (p. 67).

Illustrative of the emerging new view of science, and reminiscent of Scheffler's earlier work on educational slogans and metaphors,[3] is his comparison of epistemological metaphors (pp. 56–7) showing the shift from the ubiquitous chain (and weakest link) and foundation metaphors to interwoven cables (Peirce), threads (Wittgenstein), interconnecting girders (Russell), floating piles (Popper), and, one might add, webs (Quine and Scheffler).

The examination of Peirce's educational ideas, tracing the roots of so-called 'progressive' education to him, opens a new chapter of Peirce studies besides revealing certain inconsistencies in his views of logic as a discipline.

One can only regret the omission of any mention of Peirce's theory of symbols, particularly his tripartition of icons, indexes, and symbols which, until Goodman's *Languages of Art* (1968), dominated aesthetics and the nebulous area of semiotics, the modern origins of which are also to be found in Peirce.

James mounted the barricades just in time to foment a revolution in the philosophy of mind only to follow it up with a volley of epistemological duds under the banner of Pragmatism. Scheffler's treatment of James's theory of truth is thorough but predictable stressing the degenerate consequences of stretching the Pragmatic Maxim beyond meaning to truth. Among them: James's failure to distinguish between the mutability of truth itself and estimations of truth; and the ambiguity in his notion of satisfaction between satisfaction (verification) of a belief and the satisfaction (comfort) consequent upon a belief held. Far less cavalier are James's analyses of consciousness (the famous 'stream') and habit as reducing the conscious supervision of action. Both receive the careful assessment they deserve.

---

[1] Quine [1973], p. 3.

[2] Which, incidentally bears a striking resemblance to Popper's evolutionary speculations in his [1963], p. 47 f.

[3] Scheffler [1960].

Confronted with repeated experiences of the same thing—the same colour, smell, or pain—James warns against assuming that consciousness itself exhibits the same structure as the world, that our perceptions have the same constancy as the things they are of. 'What is got twice is the same object. . . . The sameness of the *things* is what we are concerned to ascertain; and any sensations that assure us of that will probably be considered in a rough way to be the same with each other' (quoted on p. 137). However, '. . . there is no proof that the same bodily sensation is ever got by us twice'. (Quoted on p. 139.) Rather, consciousness is 'sensibly continuous': 'The transition between the thought of one object and the thought of another is no more a break in the *thought* than a joint in a bamboo is a break in the wood. It is a part of the *consciousness* as much as a joint is a part of the *bamboo*' (quoted on p. 142).

Does this mean that contiguous thoughts or sensations are qualitatively continuous? Or conversely does the location of constancy in things amount to a denial of qualitative (phenomenal) constancy altogether? If so, then James has overstated his case and established less than he intended. On the first point, Scheffler argues to the effect that it does not follow 'that adjacent or transitional thoughts are *qualitatively* continuous simply from the fact that the sequence is unbroken' (p. 142). And on the second point that 'Phenomenal properties are . . . treated in contradictory fashion by James, who takes them sometimes as objects, sometimes as ideas. We may distinguish *physical* green from *phenomenal* green, and both of these from the *idea of* phenomenal green' (p. 139). James was of course attempting to steer away from 'domino' theories of concatenated sensations, on the one hand, and theories of relations as thought but never sensed on the other. Whatever his degree of success or failure in this, Scheffler is surely correct in saying that the attempt itself 'is of the utmost significance in fostering a more realistic conception of the conscious life' (p. 143); notwithstanding, it might be added, the latter day sensationalism of sense-data.

The core of James's view of habit is the claim that 'The more of the details of our daily life we can hand over to the effortless custody of automatism, the more our higher powers of mind will be set free for their own proper work' (quoted on p. 125). Habit is thus the realm of (mostly) unconscious, routinised action, whereas conscious attention is both critical and 'innovative'. Like most philosophers who have written on habit, James produces no account of the change of awareness involved in the shift from 'conscious' to habitual control (for we remain in some sense sufficiently 'aware' when acting out of habit to control behaviour).[1]

---

[1] A notable exception is D. C. Dennett in his [1969], esp. ch. 15 on 'Awareness and Control'. *Cf.* also M. Polanyi [1958], esp. ch. 4 on 'Skills'.

Scheffler, however, complains of the over-simplified dichotomy between 'learned performance' and 'intelligence' (p. 128) invoking Ryle's notion of 'intelligent capacities'[1] as never dispensing with judgment. 'The critical point is that in certain performances, habit and intelligence are fused through practice, so that experience teaches, but not through automization' (p. 128). Elsewhere,[2] Scheffler argues that judgment-based 'critical skills' like chess are not increased in proficiency through routinisation. But is this dichotomy between intelligence and routine any better than James's between consciousness and habit? Some matters of intelligence (in the sense of critical judgment) do appear to be routinisable; for example, recurrent forms of argument in logic, strategies of play in chess, or interpretation in the performing arts. The point is not that critical judgment is never routinised but that certain skilled activities of art, science, or games provide for an indefinite development of higher levels of critical judgment in their exercise.[3] If so, then experience may well teach through automatisation, even of aspects of intelligence.

Mead might be described as a social behaviourist with a private life smuggled into his theories. This comes about through his attempts to graft a theory of the correlative development of mind and language onto an essentially Watsonian behaviourism. One could ask whether this is behaviourism at all to which the answer is, Yes, in intention at least. In Mead's words, '. . . though it is impossible to *reduce* mind or consciousness to purely behaviouristic terms—in the sense of thus explaining it away and denying its existence as such entirely—yet it is not impossible to *explain* it in these terms, and to do so without explaining it away, or denying its existence as such, in the least' (quoted on p. 170). Scheffler duly notes the risks of such an undertaking, arguing that in fact Mead equivocates between reductive and non-reductive explanation: 'He wants to deny the existence of mind as psychical stuff but not to deny it as a function or set of functions. Yet to construe it functionally is to reduce it' (p. 171). This general difficulty is a plague throughout Mead's attempts to explain individual minds and selves as products of societal patterns of communication rather than the reverse.

The main event is Mead's four-stage theory of the functional-developmental relations among Symbolism, Mind, Self, and Society. The first, 'signal-gesture' stage is straight stimulus pairing. An involuntary (unconditioned) gesture, for example, a dog's snarl, becomes a conditioned stimulus (signal) for a second dog's conditioned response in what Mead calls a 'conversation of gestures' without any 'significants'. That is, the snarl is not being used with any intent to symbolise combativeness. This would be a special instance of what Peirce calls 'indexical' signs, established through their effects.[4] (The omission of Peirce's theory of symbols is sorely felt at this point in the book.)

Symbolisation or 'significant gesture' begins when the organism, observing the effects of its involuntary responses, deliberately uses them as conditioned stimuli so that 'the individual can indicate to himself what he indicates to others' (quoted on p. 158). Moreover, 'if somebody shakes his fist in your face you assume that he has not only a hostile attitude but that he has some idea behind it' (quoted on p. 159) in the shape of a *shared* response. The logical consequence of this is that the fist shaker should be frightened by his own gesture (perhaps he is in James-Lang fashion). But here we are invited to distinguish between the response called forth 'explicitly in the individual to whom it [the gesture] is addressed, and implicitly in the individual who makes it' (quoted on p. 160)—in other words between a response and an idea of a response.

---

[1] Ryle [1949], pp. 42–5.
[2] Scheffler [1965], p. 99.
[3] See my [1976].
[4] C. S. Peirce, 'The Icon, Index, and Symbol', in *Collected Papers*, eds. C. Hartshorne and P. Wiess (Cambridge, Harvard University Press, 1960), vol. IV.

This distinction rather defuses Scheffler's counter example to the effect that 'anger may be conveyed calmly' (p. 181) but surely it is made at the expense of any straight forwardly behavioristic notion of 'shared response'. It also renders ambiguous Mead's analysis of meaning which is defined in terms of shared response. Scheffler picks up on this latter point but in a way that appears slightly off centre. He says, 'If we translate *meaning* [in the following statement] by tendency to respond, in accord with Mead's official doctrine, we get: "Our response to the tendency to respond to what we say is constantly attached to our conversation." This result is, if not meaningless, surely very difficult to render intelligible' (p. 181). Maybe not so difficult. Among the things Mead may have in mind here is the attempt to judge the effects on others and one's own reaction to them of what one says. This is not so unintelligible as quite beside the point of accounting for the meaning of the symbol itself and represents, in the rubric of this tradition, a confusion of pragmatics and semantics.

In any event, Mead maintains that 'Where the gesture reaches that situation [of shared response] it has become what we call "language". It is now a significant symbol and it signifies a certain meaning' (quoted on p. 159). Even if sense can be made of this, Mead provides at best a set of necessary and sufficient conditions for distinguishing sign-gestures (*e.g.*, a reflexively clenched fist) from symbol-gestures (*e.g.*, the Clenched Fist of Liberation). They are not, however, sufficient to distinguish linguistic from non-linguistic symbol systems. Mead appears to drift (*via*, one suspects, the notion of 'vocal gesture') from a broad notion of 'language' more or less synonymous with symbolisation in general to a narrower conception more in conformity with the view that language is but one among other forms of symbolisation. Hence, in supplying criteria for 'symbolisation' he assumes to have accomplished the same for 'language' as well.

Mind and Self (stage three) and Society (stage four) are viewed by Mead as the individual and collective achievements, respectively, of the social process of communication. 'The internalization . . . of the external conversations of gestures . . . is the essence of thinking' (quoted on p. 160) and of Mind. Self is achieved where, in Scheffler's words, 'The minded organism becomes the object of its own reflections' (p. 161). This is manifest earliest in children's role playing and games requiring an awareness of other people's roles in them. Like Peirce, Mead rejects direct knowledge of the Self in favour of a Self indirectly known through being 'objectified' in language. Society is thus a community of selves in various states of mutual influence.

One major difficulty is how to take all this as theory. Is Mead proposing a theory of origins, of development (plus origins), or as Scheffler suggests, an 'interpretation' (pp. 175–6) without 'explanation' in terms of conditions merely necessary to societal development? A case in point is Scheffler's query whether Mead's theory offers a solution to the body-mind problem. He answers, 'It might be argued that it does not, since his theory smuggles in mental contents at the very beginning' (p. 167). In that case, perhaps Mead's problem is neither one of the origins nor of analysis but of development. In other words, he seems not so much concerned with a philosophical body-mind problem as a psychological self-society problem in which the main issue is that of determining the conditions of change and development. As a psychologist, he would naturally assume an organism in constant interaction and learning from its environment. The question then is what accounts for the observed changes throughout its life span? One might locate the difference between philosophy and social psychology in that shift of emphasis.

Dewey's Human Nature and Conduct, incidentally, is subtitled, *An Introduction to Social Psychology*; and it is well to recall that even as late as 1922, the year Dewey published this book, social psychology in America was the recent offspring of the casuistic Mental and Moral Philosophy of the previous century as well as of the experimental methods of Wundt and Watson. For practical philosophers like Mead, James and Dewey, the idea was that psychology and philosophy should go together in mutual support of a comprehensive *and applied* theory of rational thought and action.

Viewed from this standpoint, the main thrust of Scheffler's examination of Dewey is that the latter's psychological conceptions of habit, impulse, deliberation, and the like are inconsistent with his philosophical conceptions of critical inquiry and social reconstruction. En route, Scheffler also manages to brush away much of the moss of accumulated misunderstanding particularly of Dewey's social and educational opinions while raising the level of commentary throughout.

Conduct for Dewey presupposes habits, and at a more general, social level, custom imposed upon the active energy ('impulses') of the growing individual. 'In conduct the acquired is the primitive. . . . In short, the meaning of native activities is not native; it is acquired. It depends upon interaction with a matured social medium' (quoted on pp. 210–11). 'Impulse' is Dewey's substitute for 'instinctive activities' referring to specific untutored responses shaped by habit and custom. '. . . we need to know about the social conditions which have educated original activities into definite and significant dispositions before we can discuss the psychological element in society. That is the true meaning of social psychology' (quoted on p. 212). Though usually channelled by habit and custom, impulse will sometimes impell the individual to deviant, new activity where habit and custom fail. The new activity, where is it more than a howl of frustration, may be the outcome of 'deliberation' the occasion of which 'is an *excess* of preferences, not natural apathy or an absence of likings' (quoted on p. 219).

Dewey writes almost in direct reply to James's melancholy claim that '[Habit] dooms us all to fight out the battle of life upon the lines of our nuture or our early choice . . . and it is too late to begin again' (quoted on p. 124). 'More "passions", not fewer, is the answer', says Dewey. 'Rationality . . . is not a force to evoke against impulse and habit. It is the attainment of a working harmony among diverse desires' (quoted on p. 219). He also stands opposed to those contemporary philosophers of education who, like R. S. Peters, tend to see reason locked in a running battle for 'control' over emotions threatening to unseat precarious rationality.[1] On the contrary, Dewey maintains, 'The man who would intelligently cultivate intelligence will widen, not narrow, his life of strong impulses while aiming at their happy coincidence in operation' (quoted on p. 219).

---

[1] Peters [1972].

Scheffler pursues Dewey's notion of impulse in another direction, however, arguing that it commits him to a 'discontinuous' view of intelligent thought (inquiry) motivated by the breakdown of established habits. Impulse as a response to stress, conflict, or other difficulty attendant upon the failure of prior habit dissolves with the resolution of the particular problem which gave it rise. 'There is no provision in Dewey's account for putting impulse to use in revising habits that are currently functioning perfectly well and adequately' (p. 225). Hence there is no continuous flow of impulse 'for general use in advance reconstruction of potential breakdowns foreseeable only in imagination' (p. 226). This 'discontinuous conception of habit-reconstruction' (p. 236), what Scheffler also calls the 'problem-theory of reflective thought' (p. 237), conflicts with Dewey's interpretation of ends and means as requiring continuous and mutual re-evaluation, as well as his 'moral advocacy of continuous social reform based on a rational prevision of consequences' (p. 226).

The root difficulty is one Scheffler rehearsed earlier in his discussion of Peirce, namely, 'The general tendency of pragmatism . . . to interpret theory as intermediary between practical problem and practical solution' (p. 251). A secondary difficulty is the construal of the meaning (Peirce) or truth (James) of theories solely in terms of observable changes in the phenomenal world. In fact, theory has a life of its own affording 'intelligible and coherent representation[s] of fundamental natural processes' (*ibid.*), which goes far beyond 'living doubt' or defective habit as the starting points of inquiry. Scheffler's analysis of these matters is sensitive and executed with great care. There is at least one major ingredient of Dewey's philosophy, however, which does not receive the attention it ought: growth.

Dewey is well known for his eschewal of 'a fixed ideal of a remote good' (quoted on p. 223) on grounds that ultimate aims have a way of distracting us from present ailments and the good to be found in reconstructive growth. To which the choral retort has been, whither growth and development if it is not in some way directed by long range principles? In stressing the role of moral principles as hypothetical and therefore fallible guides to present activity, Dewey hardly escapes the question of how (*i.e.*, on what normative grounds) the success or failure, rightness or wrongness of such activities are to be judged. Furthermore, the 'hypothetical' nature of normative principles is surely not all of a piece with that of purely explanatory hypotheses however revisible in view of their consequences. Scheffler, in mild defence, imagines Dewey replying that 'the outcomes of growth cannot thus be restricted without substituting an uncritical restraint in advance for the operation of intelligence, thereby placing an unwarranted limitation upon the freedom of activity' (p. 247). This still is no answer to the chorus and reminds one of Moore's admonition not to confuse the questions, What is good? and What things are good? Yet the impression lingers that a better case could be made for Dewey here.

As both a philosopher of science and of education, Scheffler is eminently qualified to do justice to the pragmatists whose views of the relations among knowledge, inquiry, thought, and action tend to culminate, for better or ill, in their educational proposals. These are by no means uniform, and many salient differences among the pragmatists are most vividly represented there. In sum, the book is a major contribution to the literature on the pragmatists the full measure of which shall be taken by the piece-meal appraisals to which it will undoubtedly be subjected.

A final note on format. The book is divided into four parts corresponding to each of the four philosophers, of several chapters each. Some of the chapters are no more than two or three pages long and read more like continuous sections than separate chapters. Running the chapters together in continuous, titled sections of four basic chapters would sacrifice nothing of clarity or elegance of format and, by rough calculation, would reduce the length of the book by at least twenty pages. There is a detailed index, and Scheffler is thoughtful of the student in citing readily available editions wherever possible.

V. A. HOWARD
Institute of Education,
University of London

REFERENCES

DENNETT, D. C. [1969]: *Content and Consciousness.* London: Routledge and Kegan Paul.
GRANT, C. K. [1975]: 'Thinking Hopefully', *Times Literary Supplement,* February 14.
HOWARD, V. A. [1976]: 'Artistic Practice and Skills', in D. N. Perkins and B. Leondar (*eds.*): *Art and Cognition.* Baltimore: Johns Hopkins University Press. *Forthcoming.*
PETERS, R. S. [1972]: 'The Education of the Emotions' in R. Dearden, P. Hirst and R. Peters (*eds.*): *Education and Development of Reason.* London: Routledge and Kegan Paul.
POLANYI, M. [1958]: *Personal Knowledge.* London: Routledge and Kegan Paul.
POPPER, K. R. [1963]: *Conjectures and Refutations.* London: Routledge and Kegan Paul.
QUINE, W. V. [1973]: *The Roots of Reference.* La Salle: Open Court.
RYLE, G. [1949]: *The Concept of Mind.* London: Hutchinson.
SCHEFFLER, I. [1960]: *The Language of Education.* Springfield: Charles C. Thomas.
SCHEFFLER, I. [1965]: *The Conditions of Knowledge.* Glenview: Scott Foresman.

Brit. J. Phil. Sci. 26 (1975), 353–362 *Printed in Great Britain*

Reviews

MACKIE, J. L. [1975]: *The Cement of the Universe: A Study of Causation.* Oxford: Clarendon Press. £5.25. Pp. 239.

Mr Mackie takes the title of his book from a sentence in Hume's *Abstract* and it is his aim to examine Hume's account of causation anew so that, in his own words, 'we can take over from it whatever seems to be defensible but develop an improved account by correcting its errors and deficiencies'. The first chapter is indeed a piece of radical but illuminating exegesis of Hume, based, as Mackie explains, on the pioneer work of D. C. Stove. He argues that Hume's conclusion that causation as a relation between events is nothing but regular succession is the conclusion of a complex argument involving no fewer than fourteen distinct premisses of varying degrees of plausibility. One premiss in particular, that 'no impression of power or necessity is supplied by any observed relation or sequence' fails to distinguish between two concepts of necessity, *necessity*$_1$ which distinguishes causal from non-causal sequences and *necessity*$_2$, the supposed warrant for an a priori inference from cause to effect.

Chapter 2 deals with conditional analyses of causation. Necessity$_1$ according to Mackie contains the notions of causal priority (discussed in Ch. 7) and of counterfactual truth—the effect would not have happened had the cause not occurred. The psychological origins of our beliefs in counterfactuals lie in our experience of conditions 'where things were otherwise'. Mackie's approach here concentrates on grounds of epistemological priority on singular causal statements rather than generalisations. The third chapter on causal regularities is one of the most important and original in the book. It is based on Mill's treatment of causation, rather than Hume's more simple-minded approach and adds the concept of a causal field (of standing background conditions) and that of an *inus* condition, that is 'an *insufficient* but *non-redundant* part of an *unnecessary* but *sufficient* condition'. This gives as the typical form of a causal regularity:

> In field F, all ($AB\bar{C}$ or $DG\bar{H}$ or $JK\bar{L}$) are followed by phenomenon P, and,
>
> in F, all P are preceded by ($AB\bar{C}$ or $DG\bar{H}$ or $JK\bar{L}$).

However, causal regularities, *as we know them*, are usually incomplete. We do not know all the conjuncts, positive or negative which constitute the inus conditions, A, B, . . .

Chapter 4 on 'Kant and Transcendentalism' examines, critically and destructively, Kant's views on the alleged necessary tie between cause and effect and arguments by Bennett and Strawson designed to establish weaker quasi-Kantian conclusions. Chapter 5 discusses some conclusions of Hart and Honoré (in *Causation in the Law*) and of Ducasse, testing them in the light of Mackie's conclusions about causal regularities and conditional analyses of causation. Chapter 6 examines Russell's famous claim that the concept of causality has been outmoded by the advance of physics and replaced by that of functional connection. Mackie shows that Russell's principal arguments for this contention are unsound, and that although Russell-type causation can be seen as a limiting case of causal event sequences, it cannot replace them. He then goes on to link Mill's Method of Concomitant Variation with both functional and 'neolithic' causation in a very illuminating way.

Chapter 7 discusses the direction of causation. On a simple regularity theory, there seems to be no way of accounting for the one-way directedness that common sense and most branches of science accord to the causal relation. Mackie tries (with some success) 'to define this causal priority in terms of a contrast between fixity and unfixity, and to relate it both to the dispersal of order and to the direction of explanation'.

Chapter 8 is a long and important section on the necessity of causes. Mackie lists seven different tasks which necessity, in one or other of the senses that he distinguishes, may be expected to perform. He compares the success of regularity theories of cause with Kneale-type objective necessity theories. A test case for this is the explanation of how counterfactuals are sustained by laws of nature and not sustained by accidental regularities. Mackie tries to show (plausibly but debatably) that 'the apparent power of causal laws to sustain counterfactual conditionals was to be explained without assuming that such laws differ in form or content from accidental generalizations'. Chapter 9 argues that statistical laws can be shown to agree with Mackie's version of a regularity theory. Chapter 10 on extensionality discusses whether events or facts are to be taken as the terms in the causal relationship and under what conditions equivalent descriptions of causal terms can be substituted, *salva veritate*. To the first question, Mackie answers: 'We need then, to recognize both kinds of cause, producing causes and explanatory causes, events and facts'. A final chapter on teleology discusses the case made for the independence of final causes from efficient causes and concludes that 'we have found no objective processes which are in themselves teleological in a way that precludes their being instances of efficient causation'.

An appendix contains a revised version of Mackie's *Encyclopedia* article on Mill's methods of induction.

I have assumed that the first duty of a reviewer is to say what is in the book. But I am very conscious that a bare list of Mackie's main conclusions gives a poor impression of what is a very closely argued and important book. Only a detailed consideration of his arguments can do justice to his striking and ingenious re-statement of the case for a contemporary version of the regularity theory. Although Hume inspired the project, it is Mill rather than Hume who is the source of the account that Mackie develops. But he has added so much of his own that the book is much more than just a closely textured and ingeniously worked out set of variations on Millian themes.

He is interested throughout principally in the ontological status and nature of causes and effects and the relations between them. But he is careful to distinguish points about the psychological origins and the epistemological justification of beliefs about causes from these more basic questions.

I would have liked to see a little more about necessary and sufficient conditions, concepts which are used by Mackie (as they are by other writers) rather freely as if they were more transparent than they are. And it would have been interesting to see how the various senses of 'necessity' that are distinguished are related to the different senses of 'possibility' with which they have traditionally been entangled. But it is perhaps unreasonable to ask for further elaborations of argument in a book already so richly and subtly detailed. It is one of the virtues of Mackie's treatment that it raises all sorts of interesting peripheral questions.

The book is not an easy one to read. But this is not due to any defects of Mackie's style of writing which is clear and straightforward. It is simply that the sheer concentration of argument and examples calls for a degree of care and attention beyond what is ordinarily demanded in contemporary philosophy.

The exceptional qualities of the book are such that they make mere praise seem rather patronising. It is a singularly impressive and powerful essay and one that will provide a basis for comment and controversy about causation and allied problems for a long time to come.

D. J. O'CONNOR
University of Exeter

LANDÉ, A. [1973]: *Quantum mechanics in a new key*. New York: Exposition Press. $6.50. Pp. vii+131.

This book is little more than a second edition of Landé's *New Foundations of Quantum Mechanics* (Cambridge 1965) which was itself little more than a second edition of his *From Dualism to Unity in Quantum Theory* (Cambridge 1960). However, in the period since the publication of the earlier books, the formal obstacles to carrying through any straight-forward realist interpretation of the quantum theory have been powerfully re-emphasised in the literature, so a further analysis of Landé's particular claims seems due.

I shall restrict myself in this review to three of Landé's most central contentions. (I) he purports to have obtained a consistent particle interpretation of quantum mechanics in which the particles have the ordinary classical attributes, possessed simultaneously and independently of measurement (cf. pp. 35, 99–100, and *passim*). (II) he claims to be able to derive all the fundamental principles of quantum theory from simple plausible postulates of symmetry and the like (pp. 5, 30–73). (III) he defends a particular mechanical particle interpretation of the two slit experiment, an interpretation which purports to remove the mysteries (pp. 14–21, 74–5). In principle these three claims of Landé could stand or fall independently of each other.

Claim (I), though Landé seems never to have realised this, is in deep trouble with the no-hidden-variable results (cf. Gardner [1972], whose critical comments on Popper are, of course, equally applicable to Landé). Specifically, the theorem of Kochen and Specker (or equally the theorem of Bell, Wigner, and others) (these are all rigorous refinements of von Neumann's (1932) contentions) showed that no theory which assumes that measurements merely discover the hidden values of the measured observables can be consistent with the predictions of the quantum theory. It does not matter whether in other respects the proposed theory is deterministic or indeterministic. These theorems seem directly to rule out Landé's fundamental position.

At first it seems that Landé might find an escape route by rejecting the natural assumption that on his theory the results of measurement yield the immediately prior values of the measured observables. But (*i*) if the results of measurement depended jointly on the immediately prior values of the measured observables and on some random or non-random additional parameters, there would still be conflict with the Kochen and Specker theorem, (*ii*) if the results of measurement did not depend at all on the postulated prior values of the measured observables, these prior values would be totally unconnected with the world of appearances and any child could invent them arbitrarily: we would have a parody of a metaphysical theory with no possible epistemological bearing on the quantum theory, (*iii*) even then, as Gardner has pointed out, such a metaphysical theory could not only not agree with the actual measurement results, but could not even specify a range of possibilities consistent with the actual possibilities of measurement results. Landé's much-heralded realist particle interpretation of the quantum theory seems thus to be simply wishful thinking based on ignorance of the relevant mathematical theorems.

On the other hand, Landé's claim (II) to have given a new derivation of the quantum theory from very simple starting postulates (involving conditions on transition probabilities between states) is, as far as I can see, totally independent of claim (I) and in fact has nothing particular to do with his alleged consistent particle viewpoint. Nevertheless claim (II) has in fact occupied the bulk of all three of the books referred to.

Landé on page 41 gives a law linking transition probabilities which he calls 'the ordinary law' (pp. 41, 45) and 'the classical law' (p. 42). Landé then proceeds to show that this law cannot be true in general (even classically) and he proposes instead a law involving complex probability amplitudes which is immune to his particular objection and which yields the previous law when we carry out a certain averaging process. But the trouble is that the law Landé starts with is no classical ordinary law, but simply a fallacy. In effect he assumes that if a box is divided into eight compartments by three partitions, North–South, East–West, and top–bottom, then the probability of a particle being in the top section if in the North section must classically be *equal* to the *sum* of the probability of being in the top section if in the East section *times* the probability of being in the East section if in the North section *plus* the probability of being in the top section if in the West section *times* the probability of being in the West section if in the North section: $p(\mathrm{T,N}) = p(\mathrm{T,E})p(\mathrm{E,N})+p(\mathrm{T,W})p(\mathrm{W,N})$, where E and W are mutually exclusive and jointly exhaustive. But this law (*pace* Landé) is not classically correct: thus if all the conditional probabilities appearing on the right are set equal to $1/2$, the conditional probability on the left could classically still be anything between 0 and 1. Landé has not correctly stated any classical law; indeed the correct classical law requires more information than that provided by the various factors on the right. There is therefore no question here of an *a priori* refutation of a classical law or of an *a priori* justification for the introduction of complex probability amplitudes.

Having got his law relating complex transition probability amplitudes by this very peculiar heuristic argument, Landé attempts (pp. 67–73) to justify the quantum mechanical commutation relations on the basis of some kind of covariance requirement. His requirement is that the matrix element, of any observable which is a function of position only ('relative to an arbitrary zero point'), between two momentum states should depend only on the momentum difference between those states. There is a similar requirement with position and momentum systematically interchanged. Since Landé has thus far offered no physical interpretation of such matrix elements, the physical justification for these requirements is obscure. In particular it is not clear why he chooses position and momentum rather than, say, position and energy, or energy and momentum, as his pair in this relationship, unless it be by hindsight with respect to the results he wishes to obtain. (Landé makes no appeal to the concept of the momentum operator as the generator of infinitesimal translations, or to anything else that one might expect to play a role in this context in providing a non *ad hoc* justification of some special connection between position and momentum: as it is the connection is just pulled out of a hat). The situation is even worse when it comes to the energy-time commutation relation. Landé simply posits the same covariance relationship between energy and time, without any clear justification and in spite of the fact that time is not an observable in any straight-forward sense. Of course even both the ensuing commutation relations are still not sufficient to give Landé the time-dependent Schrödinger equation, since he has not shown that the energy can be expressed by a Hamiltonian function. In fact on page 73 Landé simply pulls the time-dependent Schrödinger equation out of a hat without any real pretence at justification.

One has to conclude that Landé's claim in his Introduction (pp. vi–vii) that he has now achieved

> a satisfactory deduction of the quantum mathematics from simple, general, and plausible non-quantal postulates, namely those of symmetry, correspondence, and covariance

fails to stand up to any remotely close scrutiny in spite of the plaudits it has received from a certain well-known philosopher.

(III), Landé's interpretation of the two slit experiment (pp. 17–21), is that the screen with the slits in reacts as a dynamical whole which is only capable of transferring momentum in discrete amounts to the electrons that pass through its slits. Which discrete amounts of momenta it can transfer depends on whether or not it has a second slit in it. It is the dynamical behaviour of the screen rather than that of the electrons which interact with it that produces the interference phenomenon and there is no mystery about the dynamical behaviour of the screen being altered when we close one slit. Landé nowhere discusses the immediate *prima facie* objection that the two slit interference pattern disappears if one puts a second screen behind one of the slits but not in contact with the first screen and placed sufficiently close to the first screen that it cannot directly interact with electrons that have passed through the other slit. This objection might be met by making use of the requirement that both screens would still need to be connected to a common rigid support. But it is not clear how the argument would work.

In mathematical justification of his interpretation, Landé cites Duane's old quantum theory interpretation of diffraction by a crystal, with Duane's rule $\Delta p_x = h/\lambda_x$, as analogous, but he does not give any mathematical details for the actual two slit case. In earlier publications Landé cited Epstein and Ehrenfest's initial (1924) attempt to apply Duane's technique to the two slit case. But in 1927 Epstein and Ehrenfest wrote a second paper pointing out in considerable mathematical detail that this approach did not in fact work for the ordinary two slit interference pattern, but only for the limiting case where the final detection screen is infinitely distant from the screen with the slits in. Landé seems now to have realised that it is not to his advantage to cite the work of these authors. Landé does argue later (pp. 74–6) that conditions equivalent to Duane's selection rules on the momentum transfer can be obtained from modern quantum theory by considering the matrix element between momentum states of the perturbing potential generated by the periodic structure. I am not convinced that an analogous treatment for the two slit system would be quantum mechanically correct. For here it is important that the final detection screen be in some kind of rigid connection with the screen with the slits in (if an interference pattern is to be produced) but in that case there is no resultant momentum transfer overall (in the direction of the line joining the slits) to the slit screen plus detection screen. To put it slightly differently, that particular piece of perturbation theory would indeed be appropriate to an experiment in which we determined the momentum distribution of the electrons after the interaction with the two slit system, but this we cannot, it seems, do without interfering with the interference pattern. The position an individual electron arrives at on the detection screen does not tell us the momentum component that was transferred to it by the slit screen because we do not know which slit it came through. It would do so only in the limiting case of an infinitely distant final detection screen. It is therefore not surprising from the point of view of the modern quantum theory that the Epstein–Ehrenfest approach to the two slit experiment should have yielded the result it did. What is surprising is that Landé should be attempting to revive it without even alluding to its apparent failure, let alone having any suggestion as to how to remove the difficulty.

JON DORLING
Chelsea College, University of London

REFERENCE

GARDNER, M. R. [1972]: 'Quantum-Theoretical Realism: Popper and Enstein *vs.* Kochen and Specker', *British Journal for the Philosophy of Science,* 23, pp. 13–23.

DINGLE, HERBERT [1972]: *Science at the Crossroads.* London: Martin Brian and O'Keeffe. Pp. 256. £3.00

Professor Dingle has been a tenacious critic of Einstein's special theory of relativity for many years. The present book gives a detailed account of his struggle to get his criticisms published and either rebutted or accepted. It is an astonishing story.

Dingle's feeling of disillusion with the theory, on which he had previously written a textbook in the Methuen Monographs on Physical Subjects series, began in 1955 when, in the course of reading Sir George Thomson's book *The Forseeable Future,* he encountered a speculation arising from the well known clock paradox of relativity, sometimes referred to as the paradox of the travelling twin. This gave rise to a controversy between him and supporters of Einstein, notably Professor W. H. McCrea. So far from being convinced by his critics, Dingle eventually came to believe that there is a fundamental contradiction in special relativity itself, irrespective of the particular difficulty he had encountered in standard accounts of the clock paradox. If Dingle's argument is accepted, special relativity must be rejected. In its place he is inclined to support a ballistic theory of the transmission of light as advocated early this century by Ritz. Dingle maintains that, contrary to general belief, Ritz's theory has never been tested, since, in his view, deductions from double star observations that were thought to rule it out are inconclusive.

The basic objection Dingle has to special relativity is that, according to the theory, two similar clocks, *A* and *B*, in uniform relative motion work at different rates, but since the relation of each to the other is symmetrical it follows that, if *A* works faster than *B*, *B* must work faster than *A*. He argues that this is impossible and so the theory must be false.

Dingle earlier published his account of this difficulty in the book *A Threefold Cord* (Allen and Unwin, 1961) that he wrote jointly with the late Viscount Samuel. He had previously failed to get his criticism published by the Royal Society, the Physical Society and *The Philosophical Magazine*. No reviewer of the book mentioned his criticism of special relativity, but in September 1962 *Nature* published a letter by him on the subject. This brought many replies but only Max Born's was separately printed in *Nature*, with Dingle's answer to it. Dingle was unimpressed by Born's conclusion that his objections to special relativity were 'just a matter of superficial formulation and confusion'.

From this point on Dingle became convinced that the failure to answer his objection was mainly due to a widespread belief that special relativity is above criticism. He felt that even more significant than his exposure of the illogical nature of special relativity was the lack of scientific integrity on the part of those who refused to respond to his argument, either by refuting it or else by accepting it. The then editor of *Nature* refused to publish his letter on the moral issue involved. So Dingle wrote to several well known Fellows of the Royal Society with the object of getting his criticism of special relativity published by that Society. Unsuccessful in this endeavour, he eventually wrote to the Editor of *Nature* a letter in which he asked the Royal Society to publish in *Nature* either a statement of the fallacy in his argument or an acknowledgment that there is no fallacy. The Editor refused to print his letter and Dingle appealed on this issue to the Press Council, but without success. Meanwhile he had circularised some 300 Fellows of the Royal Society. The upshot was that Professor W. H. McCrea agreed to write a reply to an article by Dingle setting out afresh his criticism of special relativity. Their discussion was published in *Nature* in October 1967, together with an article by the Editor stating his view of the controversy. These articles are reproduced in the Appendix of this book.

The controversy did not end with these publications, and by 1971 Dingle felt that it was necessary for him to give a full account of it in the form of a book. The main text is divided into two parts of approximately equal length. The first is devoted to the moral issue and the second to the intellectual issue. Typical of the attitude of many physicists referred to by Dingle in Part One was that taken by Lord Blackett, who said that he had taught the special theory of relativity for many years but felt unable to pronounce on the subtle point at issue, although he had found among his colleagues none who had serious doubts about the usual formulation of the theory. He refused to consider Dingle's argument. Two experts in relativity who cannot be accused of failing to consider Dingle's argument are Professor J. L. Synge and, as already mentioned, Professor McCrea. After a lengthy correspondence, Synge concluded that the contradiction described by Dingle is due to the incompatibility of special relativity with "the concept of clocks that run regularly, as understood by Professor Dingle". McCrea's rejection of Dingle's argument was based, *inter alia*, on his contention that Dingle had invoked the notion of distant simultaneity which is illegitimate in special relativity.

Dingle's argument as set out in the Appendix runs as follows. A and B are two similar clocks that 'coincide' (*i.e.* are adjacent to each other) at time $t = 0$ on clock A and $t' = 0$ on clock B. (Unprimed symbols refer to measures according to A and primed symbols to measures according to B). A and B are in uniform relative motion of velocity V. A third clock H is permanently at rest with respect to A on the line of the relative motion of A and B. The clock B 'coincides' with H at some time t'_1, on B's clock, where t'_1 is positive. A fourth clock N is permanently at rest with respect to B and 'coincides' with A at time t_2 on A's clock, where t_2 is positive. Dingle's discussion concerns two events E_1 and E_2, where E_1 denotes the instantaneous 'coincidence' of B and H, and E_2 the instantaneous 'coincidence' of A and N.

The event E_1 has space and time coordinates $(0, t_1')$ with respect to B. The corresponding coordinates of E_1 with respect to A are obtained by means of the Lorentz formulae and are (Vt_1, t_1) where

$$t_1 = \beta t_1' \tag{1}$$

β denoting the reciprocal of $\sqrt{(1-V^2/c^2)}$. Similarly, the space and time coordinates of E_2 with respect to A are $(0, t_2)$, and with respect to B they are $(-Vt_2', t_2')$, where

$$t_2' = \beta t_2. \tag{2}$$

Since clocks A and B each read zero when A and B are together and the time of occurrence of E_1 is t_1 according to A and t_1' according to B, Dingle concludes that

$$\frac{\text{rate of clock } A}{\text{rate of clock } B} = \frac{t_1}{t_1'} = \beta, \tag{3}$$

by equation (1). Similarly, since the time of occurrence of E_2 is t_2 according to A and t_2 according to B, Dingle concludes that

$$\frac{\text{rate of clock } A}{\text{rate of clock } B} = \frac{t_2}{t_2'} = \frac{1}{\beta}, \tag{4}$$

by equation (2). Since equations (3) and (4) are incompatible, Dingle maintains that special relativity must be rejected.

Dingle's argument clearly turns on the meaning that he attaches to the terms 'rate of clock A' and 'rate of clock B'. Instead of using this terminology, we can appeal to the definitions of t_1, t_1', t_2 and t_2'. In place of (3) we can write

$$\frac{\text{time according to } A \text{ of event } E_1 \text{ at } B}{\text{time according to } B \text{ of event } E_1 \text{ at } B} = \frac{t_1}{t_1'} = \beta, \tag{5}$$

and in place of (4) we can write

$$\frac{\text{time according to } B \text{ of event } E_2 \text{ at } A}{\text{time according to } A \text{ of event } E_2 \text{ at } A} = \frac{t_2'}{t_2} = \beta. \tag{6}$$

Unlike equations (3) and (4) these two equations are compatible with each other, since (6) can be obtained from (5) by interchanging A and B and by replacing E_1 by E_2, in accordance with the principle of relativity. We may therefore conclude that a contradiction arises only if we use Dingle's 'rate of clock' concept in the way he does.

The apparently contradictory results that clock A lags behind N, and hence behind B, and that clock B lags behind H, and hence behind A, are particular cases of a more general result in special relativity that can be immediately deduced from the Lorentz formulae. Consider the linear continuum of clocks on the line AB that are at rest relative to B, and the linear continuum of clocks on this line that are at rest relative to A. Choose the set of events E on this line to which all the A-clocks assign the same time t. The corresponding times t' of these events according to the B-clocks will range from indefinitely large positive values for remote events in the direction of the relative motion of A with respect to B, through the values βt on N (the B-clock instantaneously coincident with A) and t/β on B to indefinitely large negative values for remote events in the direction of the relative motion of B with respect to A. Thus the events E to which all the A-clocks assign the same time t occur at all possible times t' according to the B-clocks, only one of which (depending on t) gives a value of t' that is identical with the value of t. In general, this clock lies between A and B. On one side of it are B-clocks that appear to run more slowly than A and on the other are B-clocks that appear to go faster. A similar state of affairs prevails for A-clocks when we consider events to which all the B-clocks assign the same time t'.

Dingle's requirement that the epochs assigned to *any* event by A and B, respectively, should always be in the same ratio implies that by a new choice of time unit for one of these clocks it could be arranged that the times assigned to any given event by A and B would be the same. Dingle's requirement is therefore equivalent to introducing the Newtonian concept of universal time, and this is incompatible with special relativity.

Dingle has other reasons for rejecting special relativity. In particular, he queries the statement made by Einstein in his famous paper of 1905, that 'a balance-clock at the equator must go more slowly, by a very small amount, than a precisely similar clock situated at one of the poles under otherwise identical conditions'. Dingle asks how it is possible from Einstein's theory to conclude that the equatorial and not the polar clock works more slowly? He complains that on this question his critics have been silent. This question is a side issue compared with the problem considered above, but it is unfortunate that no one has bothered to discuss it. For a supporter of relativity, the essential difference between the two clocks is that relative to the centre of the Earth (which for the purpose concerned can be regarded as the origin of an inertial frame) the clock at the equator describes a circle and so cannot be associated with an inertial frame, whereas the polar clock is at rest and can be associated with an inertial frame for a period of time during which the curvature of the Earth's orbit can be neglected. The time difference mentioned by Einstein can be demonstrated by means of the Minkowski diagram, in which the track of the polar clock will be rectilinear whereas that of the equatorial clock will be curved. Although this answer will not be acceptable to Dingle as a valid statement in physics, because he believes that Minkowski's idea of space-time is a mathematical device that has had a calamitous effect on that subject, it is the simplest way in which the difference between the two clocks can be explained on Einstein's theory. Whether the difference exists empirically is, of course, another matter.

Even if one does not agree with Dingle's views, one can sympathise with him over the treatment he has received from some of the supporters of special relativity. For not all have been as courteous as McCrea and Synge. Moreover, different critics have often tended to give him different answers. It is not necessary to accept Dingle's objections to relativity in order to respect his courage and intellectual integrity in standing up for his beliefs. He tends, however, at times to be rather more censorious of his critics than I believe is justified. It is a pity that, although he charges some of them with being evasive, he himself is not as explicit as one would wish about the dangers that he believes will follow if physicists continue to make use of special relativity.

Like his other publications, Dingle's book is beautifully written. It also contains much interesting material on the history of relativity.

G. J. WHITROW
Imperial College

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

PROGRAMME OF MEETINGS 1975-76

1975

13 October

Professor ALAN MUSGRAVE: Logicism

10 November

Professor IRVING M. COPI: Methodologies of Frege, Russell and Wittgenstein

8 December

Professor SIR ALISTER HARDY: Behaviour as a Creative Innovator within Darwinian Evolution

1976

19 January

Professor MAURICE H. F. WILKINS: Is Philosophy of Science Relevant to Current World Problems?

16 February

Dr. M. L. G. REDHEAD: Some Philosophical Aspects of Particle Physics

8 March

4.15 p.m. ANNUAL GENERAL MEETING

5.15 p.m. Presidential Address: Professor D. J. O'CONNOR: Emergent Properties

26 April

Professor N. TINBERGEN: Man—Guinea Pig of Evolution

17 May

Professor E. C. ZEEMAN: The Qualitative Language of Mathematics

14 June

Professor RICHARD JEFFREY: Truth and Probability

24-26 September

Annual Conference at Exeter University

Meetings are held in the Haldane Room, University College, London, beginning at 5.15 p.m. Tea is served from 4.45 p.m.

Foundations of Language

International Journal of Language and Philosophy

Board of Editors: W. Abraham, Groningen; Richard D. Brecht, Harvard; Bruce Fraser, Boston University; Morris Halle, Mit; Peter Hartmann, Konstanz; K. Kunjunni Raja, Madras; Benson Mates, University of California; J. F. Staal, University of California; Pieter A. Verburg, Groningen; John W. M. Verhaar, Jakarta.

Foundations of Language deals with language and the foundations of its study, a subject which is basic to contemporary linguistics and philosophy and which is rapidly becoming indispensable to an increasing number of other disciplines. The journal appears at a time when there is increasing awareness among scholars who deal with language that attempts should be made to exchange information relating to their respective domains of investigation. Conditions favourable to such exchange are created by this journal, which neither opposes, nor adheres to, any particular trend or movement, but endeavours to publish general or technical contributions which either deal with the foundations of language, study the impact of language on various disciplines, or render the results of research in particular areas fruitful for other fields of study.

Subscription price per volume of four issues Dfl. 120,- / US $48.00, plus Dfl. 12,50 / US $5.00 for postage and handling. Six issues will be published yearly. Personal subscription price on request.

D. Reidel Publishing Company

Dordrecht - Holland / Boston - U.S.A.

NOTRE DAME JOURNAL OF FORMAL LOGIC

Bolesław Sobociński, *Editor*

Ann W. Lauer, *Business Manager*

Published quarterly by the University of Notre Dame in January, April, July, and October of each year, the Journal is devoted primarily to logic and the foundations of mathematics (i.e., algebraic logic, classical logical systems, constructive mathematics, methodology of deductive systems, model theory, non-classical formal systems, proof theory, recursion theory, and set theory).

Articles for publication and all editorial correspondence should be sent to the *Editor, P.O. Box 28, Notre Dame, Indiana 46556, U.S.A.*

Effective 1 July 1975 and beginning with Volume 17 (1976), subscriptions are $20.00 per year for institutions and $15.00 for individuals. Back volumes are $10.00 each for Volumes 1-9, $16.00 for Volumes 10-16. Orders for subscriptions and back volumes (accompanied by payment), and changes of address should be sent to the *Business Manager, P. O. Box 5, Notre Dame, Indiana 46556, U.S.A.*

Synthese

An International Journal for Epistemology and Philosophy of Science

Editor-in-Chief: Jaakko Hintikka, Academy of Finland and Stanford University.

Associate Editor: Ilkka Niiniluoto, Institute of Philosophy, Helsinki.
Review Editor: Dagfinn Føllesdal, Institute of Philosophy, University of Oslo.

Editorial Assistant: Juha Manninen, Institute of Philosophy, Helsinki.

Board of Consulting Editors: David M. Armstrong, Yehoshua Bar-Hillel, Arthur W. Burks, Robert S. Cohen, Donald Davidson, Hans Freudenthal, Ian Hacking, Jerrold J. Katz, †Imre Lakatos, G. Nuchelmans, Wesley C. Salmon, Wolfgang Stegmüller, Erik Stenius, Patrick Suppes, Klemens Szaniawski, Ladislav Tondl, H. Tornebohm, Marx W. Wartofsky.

Synthese publishes articles in all the fields covered by the subtitle. These include the theory of knowledge; the general methodological problems of science; the methodological and foundational problems of the different departmental sciences; those aspects of symbolic logic and of the foundations of mathematics which are relevant to the philosophy and methodology of science; and those facets of the history and sociology of science which are important for contemporary topical pursuits.

Subscription price per volume of four issues Dfl. 100,- / US $40.00, plus Dfl. 12,50 / US $5.00 for postage and handling. Two volumes will be published yearly. Personal subscription price on request.

D. Reidel Publishing Company
Dordrecht - Holland / Boston - U.S.A.

MIND

A Quarterly Review of Philosophy

edited by Professor **DAVID W. HAMLYN**

CENTENARY NUMBER

Vol. LXXXV — January 1976 — No. 337

Articles

Editorial: A Hundred Years of *Mind:*	THE EDITOR
George Croom Robertson: Editor 1876-1891:	A. QUINTON
G. F. Stout's Editorship of *Mind* (1892-1920):	J. A. PASSMORE
Mind under G. E. Moore (1921-1947):	C. LEWY
Gilbert Ryle's Editorship:	G. J. WARNOCK
Mind under the Editorship of David Hamlyn:	J. N. FINDLAY
Improvisation:	G. RYLE

Discussions:

Intention and Foresight in Law:	R. LYONS
Morality and Egoism:	C. H. WHITELEY
Zombies, Schizophrenics, and Purely Physical Objects:	D. LOCKE
Double Deception:	T. S. CHAMPLIN
A Rebuttal of Pollock's "Refutation" of Non-cognitivism:	F. E. MCDERMOTT
On the Recurrence of Things Past:	D. W. THEOBALD
The Justification of Deduction:	S. HAACK

Book Reviews:
Books Received:

Single issues: £1.25 — Annual Subscription: £4.00

Published for the MIND ASSOCIATION by

BASIL BLACKWELL — 49 BROAD STREET OXFORD

1 3 F[illegible] 1976

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1975-6

President: Professor D. J. O'Connor
Vice-President: Professor J. W. N. Watkins
Hon. Secretary: Helena Sheiham
Hon. Treasurer: Mr. D. Miller
Members: Mr. C. Howson, Dr M. Machover, Mr H. Martins, Mr. A. Young

EDITORIAL BOARD OF THE JOURNAL

Professor R. B. Braithwaite
Professor H. Dingle
Professor Sir John Eccles
Professor Sir Harold Jeffreys
Professor W. C. Kneale
Professor S. Körner
Professor M. Polanyi
Professor Sir Karl Popper
Professor G. J. Whitrow
Professor J. H. Woodger

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules :

' The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all persons who have been proposed by two members, approved by the Committee, and elected at an ordinary meeting. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £1·00 ($2.50). Those who wish to join should write to the Hon. Secretary, Helena Sheiham, Department of Philosophy, London School of Economics, Houghton Street, London WC2A 2AE.

Brit. J. Phil. Sci. Vol. 26 No. 4 December 1975

The British Journal for the Philosophy of Science

Contents

Articles

275–301 JERZY GIEDYMIN Antipositivism in Contemporary Philosophy of Social Science and Humanities.
303–318 GEOFFREY SAMPSON Theory Choice in a Two-level Science.
319–330 PAUL ERNEST A Critique of some Formal Theories of Meaning.

Review Articles

331–342 ERNEST GELLNER Beyond Truth and Falsehood (review of P. K. Feyerabend: *Against Method*).
343–351 V. A. HOWARD The Pragmatic Maximum (review of I. Scheffler: *Four Pragmatists*).

Reviews

353–362 D. J. O'CONNOR on J. L. Mackie: *The Cement of the Universe*; JON DORLING on A. Landé: *Quantum Mechanics in a new key*; G. J. WHITROW on H. Dingle: *Science at the Crossroads*.

Programme of Meetings

UNIVERSITY

13 FE 1976

Printed in Great Britain at the University Press, Aberdeen